ELSEVIER Journal of Computational and Applied Mathematics 58 (1995) 255–289

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Simultaneous time and chance discretization for stochastic differential equations☆

Matthias Gelbrich

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

Received 2 March 1993; revised 25 November 1993

Abstract

The paper deals with weak approximations of stochastic differential equations of Itô type, where convergence rates of the approximate solutions are shown using $E\|\cdot\|^p_{C[t_0,T]}$, $p \in [2, \infty)$. The rates can also be interpreted as rates for the L^p Wasserstein metrics, $p \in [1, \infty)$, between the distributions of exact and approximate solutions. The two approximation schemes considered are a combination of the time discretization methods of Euler and Milshtein with a chance discretization based on the invariance principle, and they work on a grid constructed to tune both discretizations.

Keywords: Stochastic differential equations; Discrete approximation; Convergence rates; Invariance principle

0. Introduction

This paper is designed to approximate the solution of a multi-dimensional stochastic differential equation (sde) of Itô type. The methods investigated here are based on the evaluation of the drift and diffusion coefficients in grid points, and they combine the time discretization of the side—as done for instance by the stochastic analogue of Euler's method—with the discretization of the stochastic input, the Wiener process. This combination of time and chance discretization is necessary for a computer simulation of the solution of the Itô sde. Another idea for discretizing such sde without using the Wiener process can be found in [18, 25] and is based on the approach of Doss [2] and Sussmann [24] who use a partial and an ordinary differential equation for constructing a pathwise solution of the sde, i.e., a pathwise convergence in the supremum norm is considered. Also convergence of time discrete schemes w.r.t. the mean square of the difference of

☆ This research is supported by Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung" of the Deutsche Forschungsgemeinschaft.

SSDI 0377-0427(94)00003-J

exact and approximate solution in the end point was investigated (e.g., [15, 18, 17]) as well as simulation strategies for functionals of the solution in the end point (e.g., [26]). A broad survey over various approximations of solutions for sde's is given in the monograph [11]. Platen [19] gives convergence orders of time discrete approximations—constructed via the stochastic Taylor expansion—w.r.t. the mean square of the supremum norm. The methods of order one and two considered there (the stochastic Euler method [14] and Milshtein's method [15]) will be the basis of the methods considered in the present paper. Together with these time discretizations we will discretize the Wiener process and estimate the distance between the distribution of the exact solution and the distributions of the approximate solutions—all solutions being referred to as random variables with values in a space of continuous functions. Janssen [9] gave results how the bounded-Lipschitz metric (see [3]) between the distributions of the value, taken in the nth grid point, of exact and approximate solutions of this type can be estimated by the accuracy of the approximation of the Wiener process. Using the approach of Doss and Sussmann, Römisch and Wakolbinger [21] show the Lipschitz dependence of the change of the solution on the change of the stochastic input process, both pathwise in the supremum norm or other distances.

Kanagawa [10] used a method derived from the stochastic Euler method by replacing the increments of the Wiener process by other "simpler" i.i.d. random variables: He uses L^p Wasserstein metrics ($p \geqslant 2$) between the distributions of exact and approximate solutions, thus achieving convergence rates. (For a broad survey on probability metrics see [20], on L^p Wasserstein metrics see, e.g., [7, 5]). We use the same metrics, but generalize the method of Kanagawa. For that we use as a basis the stochastic Euler method (E1) (proposed in [14]) and Milshtein's method (M1)′ (proposed in [15]) having order 1 and 2, respectively, with respect to the mean square of the supremum norm of the difference between exact and approximate solution. Since these methods use values of the drift and diffusion coefficients and of the Wiener process only in grid points t_k, the order 2 is optimal, as shown in [1]. The orders of (E1) and (M1)′ (w.r.t. the same metric) are proved in [19], as well as higher orders for methods using also iterated integrals of the Wiener process.

After having given the corresponding notations and definitions in Section 1, we shall prove estimates for the mean of the pth power ($p \geqslant 2$) of the supremum norm of the difference between the exact solution and the solution of (E1) and (M1)′, respectively, in Section 2. While in (E1) and (M1)′ the Wiener process is interpolating the solution between the grid points t_k, we construct the methods (E2) and (M2) by smoothening the Wiener process using piecewise linear and continuous interpolation over a finer grid consisting of points u_i^k inserted between t_k and t_{k+1}, respectively. Section 3 states an estimate for the aforementioned distance between the solutions of (E1) and (E2), and (M1)′ and (M2), respectively. All methods mentioned so far discretize only in time, i.e. only the differential equation itself. The methods (E3) and (M3) use instead of the increments of the Wiener process between neighbouring u_i^k other "simpler" i.i.d. random variables. In Section 4 an estimate w.r.t. the L^p Wasserstein metric between the distributions of the solutions of (E2) and (E3), and (M2) and (M3), respectively, is shown. Here the discretization of the Wiener process is done. Using the results of Sections 2–4, we can obtain an order w.r.t. the L^p Wasserstein metric between the distributions of the exact solution and the solutions of (E3) and (M3), respectively. This is done in Section 5, followed by a reasonable tuning of the orders achieved in Sections 2–4, which gives actually a rule for how fine the grid consisting of the points u_i^k should be chosen in dependence of the grid consisting of the points t_k. Thus, a higher order as in [10] is achieved, but also our methods need significantly less computations of the drift and diffusion coefficients of the sde.

1. Preliminaries

We consider a stochastic differential equation of Itô type written in integral form:

$$\text{(I)}\qquad x(t)-x_0=\int_{t_0}^{t} b(x(s))\,\mathrm{d}s+\int_{t_0}^{t}\sigma(x(s))\,\mathrm{d}w(s)$$

$$=\int_{t_0}^{t} b(x(s))\,\mathrm{d}s+\sum_{j=1}^{q}\int_{t_0}^{t}\sigma_j(x(s))\,\mathrm{d}w_j(s),\quad t\in[t_0,T],\ x_0\in\mathbb{R}^d,$$

where $w=(w_1,\dots,w_q)^{\mathrm{T}}$ is a q-dimensional standard Wiener process, $b\in C(\mathbb{R}^d;\mathbb{R}^d)$, and $\sigma\in C(\mathbb{R}^d;\mathscr{L}(\mathbb{R}^q;\mathbb{R}^d))$, and where $\sigma_j\in C(\mathbb{R}^d;\mathbb{R}^d)$, $j=1,\dots,q$, denote the columns of the matrix function $\sigma=(\sigma_1,\dots,\sigma_q)$.

Here and in the sequel we denote by C and C^i spaces of continuous and i times differentiable functions, respectively, and by $\mathscr{L}$ spaces of linear mappings. By $\|\cdot\|$ we shall denote the Euclidean norm on $\mathbb{R}^n$ ($n\in\mathbb{N}$) and the corresponding induced norm on a space $\mathscr{L}$. For any random variable (r.v.) ζ mapping a probability space $(\Omega,\mathscr{A},P)$ into a separable metric space (X,d) with the Borel σ-algebra $\mathscr{B}(X)$, the notation $D(\zeta)$ shall mean the distribution $P\circ\zeta^{-1}$ induced on X by ζ. $\mathscr{P}(X)$ shall be the set of all Borel probability measures on X.

The case that b and σ explicitly depend on the time t can be written in the form (I) by taking t as another component of x. A direct treatment of this case follows the same lines as in this paper and is carried out—for equidistant grids and bounded b and σ—in [4]. It allows to relax eventually required second order differentiability w.r.t. t in the present paper to first order differentiability w.r.t. t. For $p\in[1,\infty)$ we define a metric W_p on the set $\mathscr{M}_p(X):=\{\mu\in\mathscr{P}(X)\colon \int_X(d(x,\theta))^p\,\mathrm{d}\mu(x)<\infty,\ \theta\in X\}$ by

$$W_p(\mu,\nu):=\left[\inf\int_{X\times X}(d(x,y))^p\,\mathrm{d}\eta(x,y)\right]^{1/p}\quad(\mu,\nu\in\mathscr{M}_p(X)),$$

where the infimum is taken over all measures $\eta\in\mathscr{P}(X\times X)$ with marginal distributions μ and ν. W_p is called L^p Wasserstein metric and has the properties of a metric on $\mathscr{M}_p(X)$ (see [7]). With respect to these metrics the following theorem from [10] states a convergence result for a sequence of approximations to the solution x of (I) which are constructed over equidistant grids using both the stochastic Euler method and a substitution of the Wiener process increments between grid points by other i.i.d. r.v.'s (which are for instance easier to generate on a computer).

Theorem 1.1 (Kanagawa). *Let $\{\zeta^k,k=1,\dots,N\}$ a set of bounded i.i.d. q-dimensional r.v.'s with mean value 0 and covariance matrix I_q (unit matrix), with finite $(2+\delta)$th absolute moments for some $\delta\in(0,1]$, and with a quadratically integrable density. If b and σ are Lipschitz continuous, then the method*

$$\hat{y}_N(0)=x_0,$$

$$\hat{y}_N\left(\frac{k}{N}\right)=x_0+\sum_{j=1}^{k}\sigma\left(\frac{j-1}{N},\hat{y}_N\left(\frac{j-1}{N}\right)\right)\frac{\zeta^j}{\sqrt{N}}$$

$$\text{(K)} \qquad + \sum_{j=1}^{k} b\left(\frac{j-1}{N}, \hat{y}_N\left(\frac{j-1}{N}\right)\right)\frac{1}{N}, \quad k = 1, \dots, N,$$

$$\hat{y}_N(t) = \hat{y}_N\left(\frac{k-1}{N}\right) + (N \cdot t - k + 1)\left(\hat{y}_N\left(\frac{k}{N}\right) - \hat{y}_N\left(\frac{k-1}{N}\right)\right)$$

$$\textit{for } t \in \left[\frac{k-1}{N}, \frac{k}{N}\right), \quad k = 1, \dots, N,$$

converges for any $\varepsilon > \frac{1}{2}$ *and every* $p \in [2, 2 + \delta)$ *at the rate*

$$W_p(D(\hat{y}_N), D(x)) = \mathrm{O}(N^{-\delta/2(2+\delta)}(\ln N)^{\varepsilon}) \quad \textit{for } N \to \infty.$$

This idea of joint discretization w.r.t. time and chance (earlier considered also in [9], see Section 0 of the present paper) gave rise to the following construction leading to the definitions of the approximate solutions (E3) and (M3).

In the sequel we shall use, depending on our needs, the following general assumptions concerning (I):

(V1) There exists a constant $M > 0$ such that for all $j = 1, \dots, q$ and $x \in \mathbb{R}^d$,

$$\|b(x)\| \leqslant M(1 + \|x\|) \quad \text{and} \quad \|\sigma_j(x)\| \leqslant M.$$

(V2) There exists a constant $L > 0$ such that for all $j = 1, \dots, q$ and $x, y \in \mathbb{R}^d$,

$$\|b(x) - b(y)\| \leqslant L\|x - y\| \quad \text{and} \quad \|\sigma_j(x) - \sigma_j(y)\| \leqslant L\|x - y\|.$$

(V3) $b, \sigma_j \in C^2(\mathbb{R}^d; \mathbb{R}^d)$, $j = 1, \dots, q$, and there exists a constant $B > 0$ such that for all $j = 1, \dots, q$ and $x, y \in \mathbb{R}^d$,

$$\|b'(x) - b'(y)\| \leqslant B\|x - y\| \quad \text{and} \quad \|\sigma_j'(x) - \sigma_j'(y)\| \leqslant B\|x - y\|.$$

(V2)′ $b, \sigma_j \in C^2(\mathbb{R}^d; \mathbb{R}^d)$, $j = 1, \dots, q$, and there exists a constant $L > 0$ such that for all $j = 1, \dots, q$ and $x \in \mathbb{R}^d$,

$$\|b'(x)\| \leqslant L \quad \text{and} \quad \|\sigma_j'(x)\| \leqslant L.$$

(V3)′ There exists a constant $B > 0$ such that for all $j = 1, \dots, q$ and $x \in \mathbb{R}^d$

$$\sup\{b''(x)[h, k]\colon h, k \in \mathbb{R}^d,\ \|h\| \leqslant 1,\ \|k\| \leqslant 1\} \leqslant B,$$

$$\sup\{\sigma_j''(x)[h, k]\colon h, k \in \mathbb{R}^d,\ \|h\| \leqslant 1,\ \|k\| \leqslant 1\} \leqslant B.$$

(V4) $\sigma_i'\sigma_j = \sigma_j'\sigma_i$ for all $i, j = 1, \dots, q$.

Obviously, (V2)′ ∧ (V3)′ and (V2) ∧ (V3) are equivalent. (V1) and (V2) secure existence and uniqueness of the solution of (I), both in the strong sense (see [6]). The boundedness of σ_j in (V1) seems to be essential for the proof of Theorem 4.4.

The construction of the approximate solutions in Theorem 1.1 shall be generalized by considering—instead of *one* equidistant grid for both time and chance discretization—a not necessarily equidistant *coarse* grid for the time discretization and a *fine* grid, being a refinement of the former, for the chance discretization via the invariance principle which yields a lower convergence speed than the time discretization. To this end, we consider a grid class $\mathscr{G}(m, \Lambda, \alpha, \beta)$. Here let

$m:(0, T-t_0] \to [1, \infty)$ be a monotonously decreasing function and let $\Lambda, \alpha, \beta > 0$ be constants. Then each element G of $\mathscr{G}(m, \Lambda, \alpha, \beta)$ is constructed in the following way and has the following properties: G consists of two kinds of grid points:

- the time discretization points t_k, $k = 0, \dots, n$, with

$$t_0 < t_1 < \cdots < t_n = T \text{ and}$$

- the chance discretization points u_i^k, $i = 0, \dots, m_k$, $k = 0, \dots, n-1$, with

$$t_k = u_0^k < u_1^k < \cdots < u_{m_k}^k = t_{k+1},\ k = 0, \dots, n-1.$$

Hence, G is a combination of a coarse subgrid consisting of all points t_k relevant for the pure time discretization (see Section 2) and of a fine grid consisting of all points u_i^k needed for the discretization of the Wiener process (see Section 4). Denote by

$$h_k := t_{k+1} - t_k, \quad k = 0, \dots, n-1, \quad \text{and} \quad h := \max_{0 \leqslant k \leqslant n-1} h_k$$

the step sizes and the maximal step size of the coarse subgrid. Now G is required to satisfy the following assumptions:
(G1) $h \cdot n \leqslant \Lambda$ and $n \in \mathbb{N}$, $h \leqslant 1$;
(G2) $1 \leqslant m_k \leqslant m(h)^\alpha$ and $m_k \in \mathbb{N}$ for all $k = 0, \dots, n-1$;
(G3) $u_i^k - u_{i-1}^k = h_k/m_k \leqslant \beta h/m(h)$ for all $k = 0, \dots, n-1$, $i = 1, \dots, m_k$.

Here (G1) restricts the number of intervals of the coarse subgrid with given h which is bounded by 1 only for convenience (in order to write simpler upper bounds later). (G2) and (G3) say that each interval of the coarse subgrid is subdivided in an equidistant way by the points u_i^k, both the number of the subdivisions and the step size of the full grid being bounded by functions of h. As an example, it is easy to see that all equidistant grids which have also an equidistant coarse subgrid and satisfy $m_k = [m(h)]$, $k = 0, \dots, n-1$, belong to $\mathscr{G}(m, T-t_0, 1, 2)$.

For a grid G of $\mathscr{G}(m, \Lambda, \alpha, \beta)$ we define

$$[t]_G := t_k \text{ and } i_G(t) := k, \quad \text{if } t \in [t_k, t_{k+1}),\ k = 0, \dots, n-1,$$

$$[t]_G^* := u_i^k \quad \text{if } t \in [u_i^k, u_{i+1}^k),\ i = 0, \dots, m_k - 1,\ k = 0, \dots, n-1.$$

We construct the approximate solutions in (E3) and (M3) in three steps. The first step is a pure time discretization using the stochastic Euler method (E1) (see [14]) and the method (M1) corresponding to Milshtein's method (M1)′ (see [15]). Here only the coarse subgrid is involved.

$$\text{(E1)} \qquad y^{\mathrm{E}}(t) = x_0 + \int_{t_0}^{t} b(y^{\mathrm{E}}([s]_G))\,\mathrm{d}s + \sum_{j=1}^{q} \int_{t_0}^{t} \sigma_j(y^{\mathrm{E}}([s]_G))\,\mathrm{d}w_j(s) \quad \text{for all } t \in [t_0, T],$$

$$\text{(M1)} \qquad y^{\mathrm{M}}(t) = x_0 + \int_{t_0}^{t} b(y^{\mathrm{M}}([s]_G))\,\mathrm{d}s + \sum_{k=1}^{q} \left\{ \int_{t_0}^{t} \sigma_k(y^{\mathrm{M}}([s]_G))\,\mathrm{d}w_k(s) \right.$$
$$\left. + \sum_{j=1}^{q} \int_{t_0}^{t} \int_{[s]_G}^{s} (\sigma_k' \sigma_j)(y^{\mathrm{M}}([s]_G))\,\mathrm{d}w_j(u)\,\mathrm{d}w_k(s) \right\} \quad \text{for all } t \in [t_0, T].$$

If (V4) holds and $\tilde{b} := b - \frac{1}{2}\sum_{j=1}^{q} \sigma'_j \sigma_j$, (M1) is equivalent to the following method (M1)′ proposed by Milshtein [15]. This equivalence is an immediate consequence of Itô's formula.

$$\text{(M1)}' \qquad y^{\mathrm{M}}(t) = x_0 + \sum_{r=0}^{i_G(t)-1} h_r \tilde{b}(y^{\mathrm{M}}(t_r)) + \tilde{b}(y^{\mathrm{M}}([t]_G))(t - [t]_G)$$

$$+ \sum_{j=1}^{q} \left[\sum_{r=0}^{i_G(t)-1} \sigma_j(y^{\mathrm{M}}(t_r))(w_j(t_{r+1}) - w_j(t_r)) + \sigma_j(y^{\mathrm{M}}([t]_G))(w_j(t) - w_j([t]_G)) \right]$$

$$+ \frac{1}{2} \sum_{j,g=1}^{q} \left[\sum_{r=0}^{i_G(t)-1} (\sigma'_j \sigma_g)(y^{\mathrm{M}}(t_r))(w_j(t_{r+1}) - w_j(t_r))(w_g(t_{r+1}) - w_g(t_r)) \right.$$

$$\left. + (\sigma'_j \sigma_g)(y^{\mathrm{M}}([t]_G))(w_j(t) - w_j([t]_G))(w_g(t) - w_g([t]_G)) \right] \quad \text{for all } t \in [t_0, T].$$

In the second step, a continuous and piecewise linear interpolation of the trajectories in (E1) and (M1) between the points of the whole fine grid yields the methods (E2) and (M2), respectively.
(E2) $\tilde{y}^{\mathrm{E}}$ be continuous, and linear in the intervals $[u^k_{i-1}, u^k_i]$, $i = 1, \ldots, m_k$, $k = 0, \ldots, n-1$, with $\tilde{y}^{\mathrm{E}}(u^k_i) = y^{\mathrm{E}}(u^k_i)$, $i = 0, \ldots, m_k$, $k = 0, \ldots, n-1$,
(M2) $\tilde{y}^{\mathrm{M}}$ be continuous, and linear in the intervals $[u^k_{i-1}, u^k_i]$, $i = 1, \ldots, m_k$, $k = 0, \ldots, n-1$, with $\tilde{y}^{\mathrm{M}}(u^k_i) = y^{\mathrm{M}}(u^k_i)$, $i = 0, \ldots, m_k$, $k = 0, \ldots, n-1$.

In the third step, the Wiener process increments over the fine grid are replaced by other i.i.d. r.v.'s: Let $\mu \in \mathscr{P}(\mathbb{R})$ be a measure with mean value 0 and variance 1, and let

$$\{\xi^k_{js}: j = 1, \ldots, q;\ s = 1, \ldots, m_k;\ k = 0, \ldots, n-1\}$$

be a family of i.i.d. r.v.'s with distribution $D(\xi^0_{11}) = \mu$.

Then we can define the following methods (E3) and (M3) yielding continuous, and between neighbouring grid points linear, trajectories:

$$z^{\mathrm{E}}(u^0_0) = x_0,$$

$$\text{(E3)} \qquad z^{\mathrm{E}}(u^k_i) = x_0 + \sum_{r=0}^{k-1} h_r b(z^{\mathrm{E}}(t_r)) + h_k \cdot \frac{i}{m_k} b(z^{\mathrm{E}}(t_k))$$

$$+ \sum_{j=1}^{q} \left[\sum_{r=0}^{k-1} \sqrt{\frac{h_r}{m_r}} \sigma_j(z^{\mathrm{E}}(t_r)) \sum_{s=1}^{m_r} \xi^r_{js} + \sqrt{\frac{h_k}{m_k}} \sigma_j(z^{\mathrm{E}}(t_k)) \sum_{s=1}^{i} \xi^k_{js} \right]$$

for all $i = 1, \ldots, m_k$; $k = 0, \ldots, n-1$;

$$z^{\mathrm{M}}(u^0_0) = x_0, \text{ and for } \tilde{b} := b - \frac{1}{2} \sum_{j=1}^{q} \sigma'_j \sigma_j,$$

$$z^{\mathrm{M}}(u^k_i) = x_0 + \sum_{r=0}^{k-1} h_r \tilde{b}(z^{\mathrm{M}}(t_r)) + h_k \cdot \frac{i}{m_k} \tilde{b}(z^{\mathrm{M}}(t_k))$$

(M3)
$$+\sum_{j=1}^{q}\left[\sum_{r=0}^{k-1}\sqrt{\frac{h_r}{m_r}}\,\sigma_j(z^{\mathrm{M}}(t_r))\sum_{s=1}^{m_r}\xi_{js}^{r}+\sqrt{\frac{h_k}{m_k}}\,\sigma_j(z^{\mathrm{M}}(t_k))\sum_{s=1}^{i}\xi_{js}^{k}\right]$$
$$+\frac{1}{2}\sum_{j,g=1}^{q}\left[\sum_{r=0}^{k-1}\frac{h_r}{m_r}(\sigma'_j\sigma_g)(z^{\mathrm{M}}(t_r))\left(\sum_{s=1}^{m_r}\xi_{js}^{r}\right)\left(\sum_{s=1}^{m_r}\xi_{gs}^{r}\right)\right.$$
$$\left.+\frac{h_k}{m_k}(\sigma'_j\sigma_g)(z^{\mathrm{M}}(t_k))\left(\sum_{s=1}^{i}\xi_{js}^{k}\right)\left(\sum_{s=1}^{i}\xi_{gs}^{k}\right)\right]$$

for all $i = 1, \ldots, m_k$; $k = 0, \ldots, n-1$.

For this last step, the Wiener process w and the r.v.'s ξ_{ji}^k will have to be defined anew on a *common* probability space. The following three sections investigate the convergence rates w.r.t. the norm $E\sup_{t_0 \leqslant t \leqslant T}\|\cdot\|^p$ for $C([t_0, T]; \mathbb{R}^d)$-valued r.v.'s in each of the three steps.

For convenience we shall denote throughout the whole paper by K any constant depending only on p, the considered grid class, and on the data of the original sde (I). This means, K does not depend on the particular grid. Moreover, K may have different values at different occurrences.

The theorems in the sequel will be formulated for an arbitrary fixed grid G of the grid class $\mathscr{G}(m, \Lambda, \alpha, \beta)$. Therefore G fulfils (G1)–(G3) with the construction above.

2. Time discretization

The main result in this section is Theorem 2.6 which gives rates for the convergence of the approximate solutions in (E1) and (M1)′ against the solution of (I). Its proof shall use lemmas which are stated below. The first one provides the multi-dimensional Hölder inequality in both continuous and discrete shape:

Lemma 2.1 (Hölder's inequality). (a) *Let $p \in [1, \infty)$, $s < t$, and let $g : [s,t] \to \mathbb{R}^d$, $g(u) = (g_1(u), \ldots, g_d(u))^{\mathrm{T}}$ $(u \in [s,t])$, be a Borel measurable function such that $|g_i|^p$ is Lebesgue integrable over $[s,t]$ for $i = 1, \ldots, d$. Then*

$$\left\|\int_s^t g(u)\,\mathrm{d}u\right\|^p \leqslant (t-s)^{p-1}\int_s^t \|g(u)\|^p\,\mathrm{d}u.$$

(b) *Let $p \in [1, \infty)$ and $a_i \in \mathbb{R}^d$ for all $i = 1, \ldots, r$. Then*

$$\left\|\sum_{i=1}^{r} a_i\right\|^p \leqslant r^{p-1}\sum_{i=1}^{r}\|a_i\|^p.$$

Proof. (a) For $d = 1$, the assertion follows immediately from Hölder's inequality with factors g and 1, and exponents p and $p/(p-1)$, respectively.

For $d > 1$, we use the one-dimensional inequality for the functions g_i, $i = 1, \dots, d$, and $\|g\|^2$:

$$\left\| \int_s^t g(u)\,\mathrm{d}u \right\|^p = \left(\sum_{i=1}^d \left[\int_s^t g_i(u)\,\mathrm{d}u \right]^2 \right)^{p/2} \leqslant \left((t-s) \sum_{i=1}^d \int_s^t [g_i(u)]^2\,\mathrm{d}u \right)^{p/2}$$

$$= (t-s)^{p/2} \left(\int_s^t \|g(u)\|^2\,\mathrm{d}u \right)^{p/2} \leqslant (t-s)^{p-1} \int_s^t \|g(u)\|^p\,\mathrm{d}u.$$

(b) This assertion follows from (a), writing $\sum_{i=1}^r$ as $\int_0^r$ with the function $a(u) = a_i$ for $u \in (i-1, i]$, $i = 1, \dots, r$. $\square$

The next lemma will be used in the proof of Lemma 2.3.

Lemma 2.2. *Let $a_1, \dots, a_r \in \mathbb{R}$ be nonnegative and $p \in [1, \infty)$. Then*

$$\sum_{i=1}^r a_i^p \leqslant \left(\sum_{i=1}^r a_i \right)^p.$$

Proof. One may divide each a_i by $\sum_{k=1}^r a_k > 0$ which makes the inequality obvious. $\square$

The main tools for the proof of Theorem 2.6 are the multi-dimensional martingale inequalities which the following lemma contains in both continuous and discrete shape. Its proof will be a generalization of the one-dimensional case (see [8, 22], respectively).

Lemma 2.3. *Let $p \in [2, \infty)$. Then there exist constants $C_p, A_p > 0$ such that the following assertions hold:*

(a) *Let $(w(t), \mathscr{F}(t))_{t \in [\alpha, \beta]}$ be a one-dimensional standard Wiener process over the probability space $(\Omega, \mathscr{A}, P)$. Then for every function $g = (g_1, \dots, g_d) : [\alpha, \beta] \times \Omega \to \mathbb{R}^d$ with the properties*

(i) *$g(\cdot, \omega)$ is square-integrable over $[\alpha, \beta]$ for almost all $\omega \in \Omega$,*

(ii) *$g(u) = g(u, \cdot)$ is $\mathscr{F}(u)$-measurable for all $u \in [\alpha, \beta]$,*

we have

$$E \sup_{\alpha \leqslant s \leqslant t} \left\| \int_\alpha^s g(u)\,\mathrm{d}w(u) \right\|^p \leqslant d^{p/2-1} C_p E \left(\int_\alpha^t \|g(u)\|^2\,\mathrm{d}u \right)^{p/2}$$

for all $t \in [\alpha, \beta]$.

(b) *Let $(M_s, \mathscr{F}_s)_{s=0,\dots,r}$ be an $\mathbb{R}^d$-valued martingale (i.e., each component is a martingale). Then with $\Delta M_s := M_s - M_{s-1}$ we have*

$$E \sup_{0 \leqslant s \leqslant r} \|M_s\|^p \leqslant d^{p/2-1} A_p E \left(\sum_{s=1}^r \|\Delta M_s\|^2 \right)^{p/2}.$$

Proof. (a) Since for $\tilde{M}_i(t) := \int_\alpha^t g_i(u)\,\mathrm{d}w(u)$, $t \in [\alpha, \beta]$, $i = 1, \dots, d$, are quadratically integrable continuous martingales on $[\alpha, \beta]$ with quadratic variation $\langle \tilde{M}_i \rangle(t) = \int_\alpha^t [g_i(u)]^2\,\mathrm{d}u$ we have the following inequality (see [8, Ch. III, Par. 3, Theorem 3.1]):

$$E \sup_{\alpha \leqslant s \leqslant t} \left| \int_\alpha^s g_i(u)\,\mathrm{d}w(u) \right|^p \leqslant C_p E \left(\int_\alpha^t [g_i(u)]^2\,\mathrm{d}u \right)^{p/2}$$

for all $t \in [\alpha, \beta]$, $i = 1, \dots, d$, and with some constant $C_p > 0$. This yields, together with Lemma 2.1(b) and Lemma 2.2:

$$\begin{aligned} E \sup_{\alpha \leqslant s \leqslant t} \left\| \int_\alpha^s g(u)\,\mathrm{d}w(u) \right\|^p &= E \sup_{\alpha \leqslant s \leqslant t} \left[\sum_{i=1}^d \left(\int_\alpha^s g_i(u)\,\mathrm{d}w(u) \right)^2 \right]^{p/2} \\ &\leqslant d^{p/2-1} E \sup_{\alpha \leqslant s \leqslant t} \sum_{i=1}^d \left| \int_\alpha^s g_i(u)\,\mathrm{d}w(u) \right|^p \\ &\leqslant d^{p/2-1} C_p E \sum_{i=1}^d \left(\int_\alpha^t [g_i(u)]^2\,\mathrm{d}u \right)^{p/2} \\ &\leqslant d^{p/2-1} C_p E \left(\sum_{i=1}^d \int_\alpha^t [g_i(u)]^2\,\mathrm{d}u \right)^{p/2} \\ &\leqslant d^{p/2-1} C_p E \left(\int_\alpha^t \| g(u) \|^2\,\mathrm{d}u \right)^{p/2} \end{aligned}$$

for all $t \in [\alpha, \beta]$.

(b) Let $M_s = (M_s^1, \dots, M_s^d)$, $s = 0, \dots, r$, and $\Delta M_s^i := M_s^i - M_{s-1}^i$, $i = 1, \dots, d$, $s = 1, \dots, r$. Then, for $i = 1, \dots, d$, we have (see [22, Ch. VII, Par. 3, (19)]):

$$E \sup_{0 \leqslant s \leqslant r} |M_s^i|^p \leqslant A_p E \left(\sum_{s=1}^r (\Delta M_s^i)^2 \right)^{p/2}.$$

This yields, together with Lemma 2.1(b) and Lemma 2.2:

$$\begin{aligned} E \max_{0 \leqslant s \leqslant r} \| M_s \|^p &= E \sup_{0 \leqslant s \leqslant r} \left(\sum_{i=1}^d (M_s^i)^2 \right)^{p/2} \leqslant d^{p/2-1} E \max_{0 \leqslant s \leqslant r} \sum_{i=1}^d |M_s^i|^p \\ &\leqslant d^{p/2-1} A_p E \sum_{i=1}^d \left(\sum_{s=1}^r (\Delta M_s^i)^2 \right)^{p/2} \leqslant d^{p/2-1} A_p E \left(\sum_{i=1}^d \sum_{s=1}^r (\Delta M_s^i)^2 \right)^{p/2} \\ &\leqslant d^{p/2-1} A_p E \left(\sum_{s=1}^r \| \Delta M_s \|^2 \right)^{p/2}. \qquad \square \end{aligned}$$

A direct consequence of the Lemmas 2.3 and 2.1, parts (a) and (b), respectively, is the following

Corollary 2.4. *Let $p \in [2, \infty)$. Then there exist constants $C_p, A_p > 0$ such that*

(a) *under the assumptions of Lemma 2.3(a) for all $t \in [\alpha, \beta]$*

$$E \sup_{\alpha \leqslant s \leqslant t} \left\| \int_\alpha^s g(u)\,\mathrm{d}w(u) \right\|^p \leqslant [d(\beta - \alpha)]^{p/2-1} C_p \int_\alpha^t E \| g(u) \|^p\,\mathrm{d}u,$$

(b) *under the assumptions of Lemma* 2.3(b)

$$E \max_{0 \leqslant s \leqslant r} \|M_s\|^p \leqslant A_p (dr)^{p/2-1} E \sum_{s=1}^{r} \|\Delta M_s\|^p.$$

Also for Gronwall's lemma we need—besides its original shape—a discrete analogue.

Lemma 2.5 (Gronwall's lemma). (a) *Let* $f:[t_0, T] \to [0, \infty)$ *be a continuous function and* c_1, c_2 *be positive constants. If for all* $t \in [t_0, T]$

$$f(t) \leqslant c_1 + c_2 \int_{t_0}^{t} f(s)\,\mathrm{d}s$$

then

$$\sup_{t_0 \leqslant t \leqslant T} f(t) \leqslant c_1 \mathrm{e}^{c_2 (T - t_0)}.$$

(b) *Let* $a_0, \ldots, a_n$ *and* c_1, c_2 *be nonnegative real numbers. If for all* $k = 0, \ldots, n$,

$$a_k \leqslant c_1 + c_2 \frac{1}{n} \sum_{i=0}^{k-1} a_i \quad \textit{then} \ \max_{0 \leqslant i \leqslant n} a_i \leqslant c_1 \mathrm{e}^{c_2}.$$

Proof. (a) Known.

(b) Define $d_k := c_1 + c_2 (1/n) \sum_{i=0}^{k-1} a_i$ for $k = 0, \ldots, n$. Then for $k = 0, \ldots, n$ we have

$$d_k - d_{k-1} = \frac{c_2}{n} a_{k-1} \leqslant \frac{c_2}{n} d_{k-1} \text{ and thus } d_k \leqslant \left(1 + \frac{c_2}{n}\right) d_{k-1} \leqslant \left(1 + \frac{c_2}{n}\right)^k d_0 \leqslant \left(1 + \frac{c_2}{n}\right)^n c_1.$$

Since the sequence $(1 + c_2/n)^n$ is increasing and convergent to e^{c_2} for $n \to \infty$, we have $\max_{0 \leqslant i \leqslant n} a_i \leqslant \max_{0 \leqslant i \leqslant n} d_i \leqslant c_1 \mathrm{e}^{c_2}$. □

Now we can prove the following convergence result for the time discretization step. For $p = 2$ it was proved in [19]—we prove a generalization to the case $p \in [2, \infty)$, but using quite similar techniques.

Theorem 2.6. *Let* $p \in [2, \infty)$. *Then,*

(a) (V1) *and* (V2) *imply*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - y^{\mathrm{E}}(t)\|^p \leqslant K \cdot h^{p/2},$$

(b) (V1), (V2) *and* (V3) *imply*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - y^{\mathrm{M}}(t)\|^p \leqslant K \cdot h^p.$$

Proof. First, we observe the boundedness of the pth moment of the solution in (I): With Corollary 2.4(a) and Lemma 2.1(a), (b) and (V1) we get for all $t \in [t_0, T]$

$$
\begin{aligned}
&E \sup_{t_0 \leqslant s \leqslant t} \|x(s)\|^p \\
&\leqslant K\left(\|x_0\|^p + E \sup_{t_0 \leqslant s \leqslant t} (T-t_0)^{p-1} \int_{t_0}^{s} \|b(x(u))\|^p \,\mathrm{d}u + \sum_{j=1}^{q} E \sup_{t_0 \leqslant s \leqslant t} \left\| \int_{t_0}^{s} \sigma_j(x(u))\,\mathrm{d}w_j(u) \right\|^p \right) \\
&\leqslant K\left(1 + \int_{t_0}^{t} E\|b(x(u))\|^p \,\mathrm{d}u + \sum_{j=1}^{q} \int_{t_0}^{t} E\|\sigma_j(x(u))\|^p \,\mathrm{d}u\right) \\
&\leqslant K\left(1 + \int_{t_0}^{t} E \sup_{t_0 \leqslant s \leqslant u} \|x(s)\|^p \,\mathrm{d}u\right)
\end{aligned}
$$

and with Lemma 2.5(a)

$$
E \sup_{t_0 \leqslant t \leqslant T} \|x(t)\|^p \leqslant K. \tag{1}
$$

(a) Using the definitions (I) and (E1), we split the following difference for $t \in [t_0, T]$:

$$
\begin{aligned}
x(t) - y^{\mathrm{E}}(t) &= \int_{t_0}^{t} [b(x(s)) - b(x([s]_G))]\,\mathrm{d}s + \int_{t_0}^{t} [b(x([s]_G)) - b(y^{\mathrm{E}}([s]_G))]\,\mathrm{d}s \\
&\quad + \sum_{j=1}^{q} \left\{ \int_{t_0}^{t} [\sigma_j(x(s)) - \sigma_j(x([s]_G))]\,\mathrm{d}w_j(s) + \int_{t_0}^{t} [\sigma_j(x([s]_G)) - \sigma_j(y^{\mathrm{E}}([s]_G))]\,\mathrm{d}w_j(s) \right\} \\
&=: J_1(t) + J_2(t) + \sum_{j=1}^{q} \{J_{3j}(t) + J_{4j}(t)\}.
\end{aligned} \tag{2}
$$

Now for all $t \in [t_0, T]$ Lemma 2.1(a) and (V2) imply

$$
E \sup_{t_0 \leqslant r \leqslant t} \|J_1(r)\|^p \leqslant (T-t_0)^{p-1} L^p \int_{t_0}^{t} E\|x(s) - x([s]_G)\|^p \,\mathrm{d}s, \tag{3}
$$

$$
\begin{aligned}
E \sup_{t_0 \leqslant r \leqslant t} \|J_2(r)\|^p &\leqslant (T-t_0)^{p-1} L^p \int_{t_0}^{t} E\|x([s]_G) - y^{\mathrm{E}}([s]_G)\|^p \,\mathrm{d}s \\
&\leqslant K \int_{t_0}^{t} E \sup_{t_0 \leqslant u \leqslant s} \|x(u) - y^{\mathrm{E}}(u)\|^p \,\mathrm{d}s
\end{aligned} \tag{4}
$$

as well as Corollary 2.4(a) and (V2) imply

$$
E \sup_{t_0 \leqslant r \leqslant t} \|J_{3j}(r)\|^p \leqslant K \int_{t_0}^{t} E\|x(s) - x([s]_G)\|^p \,\mathrm{d}s, \tag{5}
$$

$$
\begin{aligned}
E \sup_{t_0 \leqslant r \leqslant t} \|J_{4j}(r)\|^p &\leqslant K \int_{t_0}^{t} E\|x([s]_G) - y^{\mathrm{E}}([s]_G)\|^p \,\mathrm{d}s \\
&\leqslant K \int_{t_0}^{t} E \sup_{t_0 \leqslant u \leqslant s} \|x(u) - y^{\mathrm{E}}(u)\|^p \,\mathrm{d}s.
\end{aligned} \tag{6}
$$

Here, with Lemma 2.1(b), (a), Corollary 2.4(a), (V1) and (1) it holds for all $s \in [t_0, T]$ that

$$\begin{aligned} E\|x(s) - x([s]_G)\|^p &\leqslant K \cdot E\left\{\left\|\int_{[s]_G}^s b(x(u))\,\mathrm{d}u\right\|^p + \sum_{j=1}^q \left\|\int_{[s]_G}^s \sigma_j(x(u))\,\mathrm{d}w_j(u)\right\|^p\right\} \\ &\leqslant K\left\{h^{p-1}\int_{[s]_G}^s E\|b(x(u))\|^p\,\mathrm{d}u + \sum_{j=1}^q h^{p/2-1}\int_{[s]_G}^s E\|\sigma_j(x(u))\|^p\,\mathrm{d}u\right\} \\ &\leqslant K\left\{h^{p/2}\left(1 + E\sup_{t_0 \leqslant t \leqslant T}\|x(t)\|^p\right)\right\} \leqslant K \cdot h^{p/2}. \end{aligned} \tag{7}$$

Summarizing (2)–(7), we get for all $t \in [t_0, T]$ that

$$\begin{aligned} E\sup_{t_0 \leqslant r \leqslant t}\|x(r) - y^{\mathrm{E}}(r)\|^p &\leqslant K\Bigg[E\sup_{t_0 \leqslant r \leqslant t}\|J_1(r)\|^p + E\sup_{t_0 \leqslant r \leqslant t}\|J_2(r)\|^p \\ &\qquad + \sum_{j=1}^q\left\{E\sup_{t_0 \leqslant r \leqslant t}\|J_{3j}(r)\|^p + E\sup_{t_0 \leqslant r \leqslant t}\|J_{4j}(r)\|^p\right\}\Bigg] \\ &\leqslant K\left\{\int_{t_0}^t E\sup_{t_0 \leqslant u \leqslant s}\|x(u) - y^{\mathrm{E}}(u)\|^p\,\mathrm{d}s + h^{p/2}\right\}, \end{aligned}$$

and with Lemma 2.5(a) the assertion follows.

(b) Using Itô's formula for b and σ_k, $k = 1, \ldots, q$, we obtain for $t \in [t_0, T]$

$$\begin{aligned} x(t) - x_0 &= \int_{t_0}^t b(x(s))\,\mathrm{d}s + \sum_{k=1}^q \int_{t_0}^t \sigma_k(x(s))\,\mathrm{d}w_k(s) \\ &= \int_{t_0}^t b(x([s]_G))\,\mathrm{d}s + \int_{t_0}^t \int_{[s]_G}^s \left((b'b)(x(u)) + \frac{1}{2}\sum_{j=1}^q b''[\sigma_j, \sigma_j](x(u))\right)\mathrm{d}u\,\mathrm{d}s \\ &\quad + \sum_{j=1}^q \int_{t_0}^t \int_{[s]_G}^s (b'\sigma_j)(x(u))\,\mathrm{d}w_j(u)\,\mathrm{d}s \\ &\quad + \sum_{k=1}^q \Bigg\{\int_{t_0}^t \sigma_k(x([s]_G))\,\mathrm{d}w_k(s) \\ &\qquad + \int_{t_0}^t \int_{[s]_G}^s \left((\sigma_k' b)(x(u)) + \frac{1}{2}\sum_{j=1}^q \sigma_k''[\sigma_j, \sigma_j](x(u))\right)\mathrm{d}u\,\mathrm{d}w_k(s) \\ &\qquad + \sum_{j=1}^q \int_{t_0}^t \int_{[s]_G}^s (\sigma_k'\sigma_j)(x(u))\,\mathrm{d}w_j(u)\,\mathrm{d}w_k(s)\Bigg\}. \end{aligned} \tag{8}$$

With the notations

$$f_b := b'b + \frac{1}{2}\sum_{j=1}^q b''[\sigma_j, \sigma_j] : \mathbb{R}^d \to \mathbb{R}^d,$$

$$f_{\sigma_k} := \sigma_k' b + \frac{1}{2}\sum_{j=1}^q \sigma_k''[\sigma_j, \sigma_j] : \mathbb{R}^d \to \mathbb{R}^d, \quad k = 1, \ldots, q,$$

and with (M1) and (8) we have for $t\in[t_0,T]$

$$
\begin{aligned}
x(t)-y^{\mathrm{M}}(t) = {} & \int_{t_0}^{t}[b(x([s]_G))-b(y^{\mathrm{M}}([s]_G))]\,\mathrm{d}s \\
& + \int_{t_0}^{t}\int_{[s]_G}^{s} f_b(x(u))\,\mathrm{d}u\,\mathrm{d}s + \sum_{j=1}^{q}\int_{t_0}^{t}\int_{[s]_G}^{s}(b'\sigma_j)(x(u))\,\mathrm{d}w_j(u)\,\mathrm{d}s \\
& + \sum_{k=1}^{q}\Bigg\{\int_{t_0}^{t}[\sigma_k(x([s]_G))-\sigma_k(y^{\mathrm{M}}([s]_G))]\,\mathrm{d}w_k(s) + \int_{t_0}^{t}\int_{[s]_G}^{s} f_{\sigma_k}(x(u))\,\mathrm{d}u\,\mathrm{d}w_k(s) \\
& \qquad + \sum_{j=1}^{q}\int_{t_0}^{t}\int_{[s]_G}^{s}[(\sigma_k'\sigma_j)(x(u))-(\sigma_k'\sigma_j)(y^{\mathrm{M}}([u]_G))]\,\mathrm{d}w_j(u)\,\mathrm{d}w_k(s)\Bigg\} \\
=: {} & I_1(t)+I_2(t)+\sum_{j=1}^{q}I_{3j}(t)+\sum_{k=1}^{q}\Bigg\{I_{4k}(t)+I_{5k}(t)+\sum_{j=1}^{q}I_{6kj}(t)\Bigg\}.
\end{aligned}
\tag{9}
$$

Similarly to (4) and (6), it holds that for all $t\in[t_0,T]$

$$
E\sup_{t_0\leqslant r\leqslant t}\|I_1(r)\|^p + \sum_{k=1}^{q}E\sup_{t_0\leqslant r\leqslant t}\|I_{4k}(r)\|^p \leqslant K\int_{t_0}^{t}E\sup_{t_0\leqslant u\leqslant s}\|x(u)-y^{\mathrm{M}}(u)\|^p\,\mathrm{d}s. \tag{10}
$$

The next estimates follow with Lemma 2.1, (V1)–(V3) and (1) for all $u\in[t_0,T]$:

$$
\begin{aligned}
E\|f_b(x(u))\|^p & \leqslant K\Bigg\{E\|(b'b)(x(u))\|^p + \frac{1}{2^p}\sum_{j=1}^{q}E\|b''[\sigma_j,\sigma_j](x(u))\|^p\Bigg\} \\
& \leqslant K\Bigg\{L^pM^pE(1+\|x(u)\|)^p + \frac{q}{2^p}B^pM^{2p}\Bigg\} \\
& \leqslant K\{1+E\|x(u)\|^p\} \leqslant K
\end{aligned}
$$

and, thus,

$$
E\sup_{t_0\leqslant r\leqslant t}\|I_2(r)\|^p \leqslant (T-t_0)^{p-1}h^{p-1}\int_{t_0}^{t}\int_{[s]_G}^{s}E\|f_b(x(u))\|^p\,\mathrm{d}u\,\mathrm{d}s \leqslant K\cdot h^p \tag{11}
$$

for $t\in[t_0,T]$. Completely the same way leads for all $t\in[t_0,T]$ via

$$
E\|f_{\sigma_k}(x(u))\|^p \leqslant K, \tag{12}
$$

using Corollary 2.4(a) and Lemma 2.1(a), to

$$
E\sup_{t_0\leqslant r\leqslant t}\|I_{5k}(r)\|^p \leqslant K\cdot h^{p-1}\int_{t_0}^{t}\int_{[s]_G}^{s}E\|f_{\sigma_k}(x(u))\|^p\,\mathrm{d}u\,\mathrm{d}s \leqslant K\cdot h^p \tag{13}
$$

for all $t\in[t_0,T]$ and $k=1,\dots,q$. Defining $g_j(u):=(b'\sigma_j)(x(u))$, $j=1,\dots,q$, we get

$$
E\sup_{t_0\leqslant r\leqslant t}\|I_{3j}(r)\|^p = E\sup_{t_0\leqslant r\leqslant t}\Bigg\|\int_{t_0}^{[r]_G}\int_{[s]_G}^{s}g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s + \int_{[r]_G}^{r}\int_{[r]_G}^{s}g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\Bigg\|^p
$$

$$\leqslant K\left\{E\max_{t_i\leqslant t}\left\|\int_{t_0}^{t_i}\int_{[s]_G}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p\right.$$

$$\left.+\,E\max_{T\neq t_i\leqslant t}\sup_{t_i\leqslant r\leqslant t_{i+1}}\left\|\int_{t_i}^{r}\int_{t_i}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p\right\}$$

$$\leqslant K\left\{E\max_{0\leqslant i\leqslant n-1}\left\|\int_{t_0}^{t_i}\int_{[s]_G}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p\right.$$

$$\left.+\sum_{i=0}^{n-1}E\sup_{t_i\leqslant r\leqslant t_{i+1}}\left\|\int_{t_i}^{r}\int_{t_i}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p\right\} \tag{14}$$

for all $j = 1,\dots,q$; $t\in[t_0, T]$. Obviously, by

$$S_i^j := \int_{t_0}^{t_i}\int_{[s]_G}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s,\quad i = 0,\dots,n;\ j = 1,\dots,q,$$

a martingale w.r.t. the filtration $\mathscr{F}_i := \mathscr{F}(t_i)$, $i = 0,\dots,n$, is defined, where $\mathscr{F}(s)$ is the σ-algebra generated by $\{w(u), u\leqslant s\}$. Indeed, we have

$$E(S_i^j\,|\,\mathscr{F}_{i-1}) = \int_{t_0}^{t_{i-1}}\int_{[s]_G}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s + \int_{t_{i-1}}^{t_i} E\left(\int_{t_{i-1}}^{s} g_j(u)\,\mathrm{d}w_j(u)\,|\,\mathscr{F}_{i-1}\right)\mathrm{d}s = S_{i-1}^j,$$

since $M_s := \int_{t_0}^{s} g_j(u)\,\mathrm{d}w_j(u)$ is an $\mathscr{F}(s)$-martingale. Then Corollary 2.4(b) gives us

$$E\max_{0\leqslant i\leqslant n-1}\|S_i^j\|^p \leqslant K\cdot n^{p/2-1}\sum_{i=1}^{n} E\|S_i^j - S_{i-1}^j\|^p. \tag{15}$$

Here, because of Lemma 2.1(a), Corollary 2.4(a), (V2) and (V1), it follows that for $i = 1,\dots,n$; $j = 1,\dots,q$,

$$E\|S_i^j - S_{i-1}^j\|^p = E\left\|\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p \leqslant h^{p-1}E\int_{t_{i-1}}^{t_i}\left\|\int_{t_{i-1}}^{s} g_j(u)\,\mathrm{d}w_j(u)\right\|^p\mathrm{d}s$$

$$\leqslant K\cdot h^{p+p/2-2}E\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\|(b'\sigma_j)(x(u))\|^p\,\mathrm{d}u\,\mathrm{d}s$$

$$\leqslant K\cdot h^{3p/2-2}L^pM^ph^2 \leqslant K\cdot h^{3p/2}. \tag{16}$$

Analogously, for $i = 0,\dots,n-1$ an estimate for the last term in (14) follows:

$$E\sup_{t_i\leqslant r\leqslant t_{i+1}}\left\|\int_{t_i}^{r}\int_{t_i}^{s} g_j(u)\,\mathrm{d}w_j(u)\,\mathrm{d}s\right\|^p \leqslant h^{p-1}E\int_{t_i}^{t_{i+1}}\left\|\int_{t_i}^{s} g_j(u)\,\mathrm{d}w_j(u)\right\|^p\mathrm{d}s$$

$$\leqslant K\cdot h^{3p/2}. \tag{17}$$

Summarizing (14)–(17), with (G1) we have for $j = 1, \dots, q$; $t \in [t_0, T]$,

$$\begin{aligned} E \sup_{t_0 \leqslant r \leqslant t} \| I_{3j}(r) \|^p &\leqslant K \left\{ n^{p/2-1} \sum_{i=1}^{n} h^{3p/2} + \sum_{i=0}^{n-1} h^{3p/2} \right\} \\ &\leqslant K\{n^{p/2} h^{p/2} + n h^{p/2}\} h^p \\ &\leqslant K\{\Lambda^{p/2} + \Lambda (T - t_0)^{p/2-1}\} h^p \leqslant K \cdot h^p. \end{aligned} \tag{18}$$

We get with Corollary 2.4(a) for k, $j = 1, \dots, q$; $t \in [t_0, T]$, using the notation $c_{kj}(u) := (\sigma'_k \sigma_j)(x(u)) - (\sigma'_k \sigma_j)(y^{\mathrm{M}}([u]_G))$ $(u \in [t_0, T])$,

$$\begin{aligned} E \sup_{t_0 \leqslant r \leqslant t} \| I_{6kj}(r) \|^p &\leqslant K \cdot E \int_{t_0}^{t} \left\| \int_{[s]_G}^{s} c_{kj}(u) \, \mathrm{d}w_j(u) \right\|^p \mathrm{d}s \\ &\leqslant K \cdot h^{p/2-1} \int_{t_0}^{t} \int_{[s]_G}^{s} E \| c_{kj}(u) \|^p \, \mathrm{d}u \, \mathrm{d}s. \end{aligned} \tag{19}$$

Here, with Lemma 2.1(b), (V1) and (V2)—and, thus, the Lipschitz continuity of $\sigma'_k \sigma_j$ with the constant $BM + L^2$— and (7) we get that

$$\begin{aligned} E \| c_{kj}(u) \|^p &\leqslant K\{E \| (\sigma'_k \sigma_j)(x(u)) - (\sigma'_k \sigma_j)(x([u]_G)) \|^p \\ &\qquad + E \| (\sigma'_k \sigma_j)(x([u]_G)) - (\sigma'_k \sigma_j)(y^{\mathrm{M}}([u]_G)) \|^p\} \\ &\leqslant K(BM + L^2)\, \{E \| x(u) - x([u]_G) \|^p + E \| x([u]_G) - y^{\mathrm{M}}([u]_G) \|^p\} \\ &\leqslant K\{h^{p/2} + E \| x([u]_G) - y^{\mathrm{M}}([u]_G) \|^p\} \end{aligned} \tag{20}$$

for k, $j = 1, \dots, q$, $u \in [t_0, T]$. (19) and (20) imply

$$E \sup_{t_0 \leqslant r \leqslant t} \| I_{6kj}(r) \|^p \leqslant K \left\{ h^{p/2} \int_{t_0}^{t} E \sup_{t_0 \leqslant u \leqslant s} \| x(u) - y^{\mathrm{M}}(u) \|^p \, \mathrm{d}s + h^p \right\} \tag{21}$$

for k, $j = 1, \dots, q$, $t \in [t_0, T]$.

Summarizing (9)–(11), (13), (18) and (21), we finally get

$$E \sup_{t_0 \leqslant r \leqslant t} \| x(r) - y^{\mathrm{M}}(r) \|^p \leqslant K \left\{ \int_{t_0}^{t} E \sup_{t_0 \leqslant u \leqslant s} \| x(u) - y^{\mathrm{M}}(u) \|^p \, \mathrm{d}s + h^p \right\},$$

and with Gronwall's lemma (Lemma 2.5(a)) assertion (b) follows. □

3. Refined time discretization with interpolation

Here, the solutions in (E1) and (M1)′—which behave like the Wiener process between two neighbouring points t_{k-1} and t_k of the coarse subgrid of G—are smoothened by linear interpolation with vertices in *all* grid points of G, that means in all u_i^k. This will be the contents of Theorem 3.4, and for its proof we need the following three lemmas.

Lemma 3.1. *Let v_i, $i = 1, \dots, r$, be i.i.d. standard-normally distributed real-valued r.v.'s. Then for all $p \in [2, \infty)$*

$$E \max_{1 \leqslant i \leqslant r} |v_i|^p \leqslant K(1 + \ln r)^{p/2}.$$

Proof. Obviously, the proof can be restricted to the case $r \geqslant 2$. Let $\chi\{H\}$ take the values 1 and 0 if the statement H is true or false, respectively. Then we have

$$\begin{aligned}
E \max_{1 \leqslant i \leqslant r} |v_i|^p &= E\left(\chi\left\{\max_{1 \leqslant i \leqslant r} |v_i|^p \leqslant 4^p(\ln r)^{p/2}\right\} \max_{1 \leqslant i \leqslant r} |v_i|^p\right) \\
&\quad + E\left(\chi\left\{\max_{1 \leqslant i \leqslant r} |v_i|^p > 4^p(\ln r)^{p/2}\right\} \max_{1 \leqslant i \leqslant r} |v_i|^p\right) \\
&\leqslant 4^p(\ln r)^{p/2} + \sum_{j=1}^{r} E\left(\chi\left\{|v_j|^p = \max_{1 \leqslant i \leqslant r} |v_i|^p, |v_j|^p > 4^p(\ln r)^{p/2}\right\} |v_j|^p\right) \\
&\leqslant 4^p(\ln r)^{p/2} + \sum_{j=1}^{r} E(\chi\{|v_j|^p > 4^p(\ln r)^{p/2}\} |v_j|^p) \\
&= 4^p(\ln r)^{p/2} + rE(\chi\{|v_1|^p > 4^p(\ln r)^{p/2}\} |v_1|^p) \\
&= 4^p(\ln r)^{p/2} + r\sqrt{2/\pi} \int_{4\sqrt{\ln r}}^{\infty} x^p \exp(-x^2/2)\, \mathrm{d}x \\
&= 4^p(\ln r)^{p/2} \left[1 + r\sqrt{8 \ln r/\pi} \int_1^{\infty} y^{(p-1)/2} r^{-8y}\, \mathrm{d}y\right],
\end{aligned}$$

after substituting $x = 4\sqrt{\ln r}\sqrt{y}$. Now it suffices to show that the term in the brackets has an upper bound independent of r. Indeed, using $r^{-3y/2} \leqslant r^{-3/2}$ and $r^{-13y/2} < \exp(-y)$ for $y \geqslant 1$ and $r \geqslant 2$, we get

$$\begin{aligned}
r\sqrt{\ln r} \int_1^{\infty} y^{(p-1)/2} r^{-8y}\, \mathrm{d}y &\leqslant \sqrt{\frac{\ln r}{r}} \int_1^{\infty} y^{(p-1)/2} r^{-13y/2}\, \mathrm{d}y \leqslant \sqrt{\frac{\ln r}{r}} \int_1^{\infty} y^{(p-1)/2} \exp(-y)\, \mathrm{d}y \\
&\leqslant \sqrt{\frac{\ln r}{r}}\, \Gamma\left(\frac{p+1}{2}\right) \leqslant \sqrt{\frac{1}{\mathrm{e}}}\, \Gamma\left(\frac{p+1}{2}\right),
\end{aligned}$$

and the lemma is proved. □

Lemma 3.2. *Let $(\tilde{w}(t))_{t \in [\tau_0, \infty]}$ be a one-dimensional standard Wiener process and x a standard-normally distributed random variable. Then, for $\tau_0 \leqslant a < \bar{a} < \infty$, the random variables $\sqrt{1/(\bar{a} - a)}\, \sup_{a \leqslant t \leqslant \bar{a}} (\tilde{w}(t) - \tilde{w}(a))$ and $|x|$ have the same distribution.*

Proof. This assertion follows with the scale and shift invariance properties of the Wiener process and a corollary to Lévy's formula (see [6, p. 352]). □

Lemma 3.3. *Let $a_0 < a_1 < \cdots < a_r$ be a partition of $[a_0, a_r]$ with maximal step size $\Delta := \max_{0 \leqslant i \leqslant r-1} (a_{i+1} - a_i)$ and $(\tilde{w}(t))_{t \in [a_0, a_r]}$ a one-dimensional standard Wiener process. Then for $p \in [2, \infty)$,*

$$E \max_{0 \leqslant i \leqslant r-1} \sup_{a_i \leqslant t \leqslant a_{i+1}} |\tilde{w}(t) - \tilde{w}(a_i)|^p \leqslant K \cdot \Delta^{p/2} (1 + \ln r)^{p/2}.$$

Proof. For $i = 0, \ldots, r-1$ it is obvious that

$$\sup_{a_i \leqslant t \leqslant a_{i+1}} |\tilde{w}(t) - \tilde{w}(a_i)|^p \leqslant \Big(\sup_{a_i \leqslant t \leqslant a_{i+1}} (\tilde{w}(t) - \tilde{w}(a_i)) \Big)^p + \Big(\sup_{a_i \leqslant t \leqslant a_{i+1}} (\tilde{w}(a_i) - \tilde{w}(t)) \Big)^p.$$

Then, since both suprema on the right-hand side have identical measures because of the invariance of the Wiener measure w.r.t. inversion of the trajectories at the time axis, the assertion follows immediately with Lemmas 3.1 and 3.2. □

Now upper bounds for the L^p-norm of the differences between the approximate solutions in (E1) and (E2), and in (M1)′ and (M2), respectively, can be obtained.

Theorem 3.4. *Let $p \in [2, \infty)$. Then*
(a) (V1) *and* (V2) *imply*

$$E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{E}}(t) - \tilde{y}^{\mathrm{E}}(t) \|^p \leqslant K \left(\frac{h}{m(h)} \right)^{p/2} \left(1 + \ln \left(\frac{m(h)}{h} \right) \right)^{p/2},$$

(b) (V1)–(V4) *imply*

$$E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{M}}(t) - \tilde{y}^{\mathrm{M}}(t) \|^p \leqslant K \left(\frac{h}{m(h)} \right)^{p/2} \left(1 + \ln \left(\frac{m(h)}{h} \right) \right)^{p/2}.$$

Proof. (a) First we consider the process $\bar{y}^{\mathrm{E}}$ with $\bar{y}^{\mathrm{E}}(t_0) = x_0$, $\bar{y}^{\mathrm{E}}(u_i^k) = \tilde{y}^{\mathrm{E}}(u_i^k)$, $\bar{y}^{\mathrm{E}}(t) = \bar{y}^{\mathrm{E}}(u_{i-1}^k)$ for $t \in [u_{i-1}^k, u_i^k)$ $(k = 0, \ldots, n-1;\ i = 1, \ldots, m_k)$. Then, with Lemma 2.1(b), (V1), Lemma 3.3, (G2) and (G3), we have

$$\begin{aligned}
& E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{E}}(t) - \bar{y}^{\mathrm{E}}(t) \|^p \\
& \leqslant K \left\{ E \sup_{t_0 \leqslant t \leqslant T} \left\| \int_{[t]_G^*}^t b(y^{\mathrm{E}}([t]_G)) \, \mathrm{d}s \right\|^p + \sum_{j=1}^q E \sup_{t_0 \leqslant t \leqslant T} \left\| \sigma_j(y^{\mathrm{E}}([t]_G)) \int_{[t]_G^*}^t \mathrm{d}w_j(s) \right\|^p \right\} \\
& \leqslant K \left\{ E \sup_{t_0 \leqslant t \leqslant T} [(t - [t]_G^*)^p M^p (1 + \| y^{\mathrm{E}}([t]_G) \|^p)] \right. \\
& \qquad \left. + M^p \sum_{j=1}^q E \max_{\substack{0 \leqslant k \leqslant n-1 \\ 0 \leqslant i \leqslant m_k - 1}} \sup_{u_i^k \leqslant t \leqslant u_{i+1}^k} |w_j(t) - w_j(u_i^k)|^p \right\}
\end{aligned}$$

$$\leqslant K\left\{\max_{0\leqslant k\leqslant n-1}\left(\frac{h_k}{m_k}\right)^p\left(1+E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)\|^p\right)\right.$$
$$\left.+\sum_{j=1}^{q}\max_{0\leqslant k\leqslant n-1}\left(\frac{h_k}{m_k}\right)^{p/2}(1+\ln(n\cdot m(h)^{\alpha}))^{p/2}\right\}$$
$$\leqslant K\left\{1+E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)\|^p\right\}\left(\frac{h}{m(h)}\right)^{p/2}(1+\ln n+\ln m(h))^{p/2}. \tag{22}$$

Since we have by Minkovski's inequality that

$$\left(E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)\|^p\right)^{1/p}\leqslant\left(E\sup_{t_0\leqslant t\leqslant T}\|x(t)-y^{\mathrm{E}}(t)\|^p\right)^{1/p}+\left(E\sup_{t_0\leqslant t\leqslant T}\|x(t)\|^p\right)^{1/p},$$

where the right-hand side is bounded because of Theorem 2.6 and (1), it holds that

$$E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)\|^p\leqslant K. \tag{23}$$

Hence, by (22) and (G1),

$$E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)-\bar{y}^{\mathrm{E}}(t)\|^p\leqslant K\left(\frac{h}{m(h)}\right)^{p/2}(1+\ln n+\ln m(h))^{p/2}$$
$$\leqslant K\left(\frac{h}{m(h)}\right)^{p/2}\left(1+\ln\left(\frac{m(h)}{h}\right)\right)^{p/2}. \tag{24}$$

On the other hand,

$$E\sup_{t_0\leqslant t\leqslant T}\|\bar{y}^{\mathrm{E}}(t)-\tilde{y}^{\mathrm{E}}(t)\|^p$$
$$=E\max_{\substack{0\leqslant k\leqslant n-1\\0\leqslant i\leqslant m_k-1}}\sup_{u_i^k\leqslant t\leqslant u_{i+1}^k}\|\bar{y}^{\mathrm{E}}(t)-\tilde{y}^{\mathrm{E}}(t)\|^p=E\max_{\substack{0\leqslant k\leqslant n-1\\0\leqslant i\leqslant m_k-1}}\|\tilde{y}^{\mathrm{E}}(u_{i+1}^k)-\tilde{y}^{\mathrm{E}}(u_i^k)\|^p$$
$$\leqslant E\max_{\substack{0\leqslant k\leqslant n-1\\0\leqslant i\leqslant m_k-1}}\sup_{u_i^k\leqslant t\leqslant u_{i+1}^k}\|y^{\mathrm{E}}(t)-y^{\mathrm{E}}(u_i^k)\|^p=E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)-\bar{y}^{\mathrm{E}}(t)\|^p. \tag{25}$$

Now, with (24) and (25) we have

$$E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)-\tilde{y}^{\mathrm{E}}(t)\|^p\leqslant K\left\{E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)-\bar{y}^{\mathrm{E}}(t)\|^p+E\sup_{t_0\leqslant t\leqslant T}\|\bar{y}^{\mathrm{E}}(t)-\tilde{y}^{\mathrm{E}}(t)\|^p\right\}$$
$$\leqslant K\cdot E\sup_{t_0\leqslant t\leqslant T}\|y^{\mathrm{E}}(t)-\bar{y}^{\mathrm{E}}(t)\|^p$$
$$\leqslant K\left(\frac{h}{m(h)}\right)^{p/2}\left(1+\ln\left(\frac{m(h)}{h}\right)\right)^{p/2}.$$

(b) As in (a), we first consider the process $\bar{y}^{\mathrm{M}}$ defined by

$$\bar{y}^{\mathrm{M}}(t_0)=x_0,\quad \bar{y}^{\mathrm{M}}(u_i^k)=\tilde{y}^{\mathrm{M}}(u_i^k),\quad \bar{y}^{\mathrm{M}}(t)=\bar{y}^{\mathrm{M}}(u_{i-1}^k)$$

for $t \in [u^k_{i-1}, u^k_i)$ $(k = 0, \ldots, n-1;\ i = 1, \ldots, m_k)$, and with

$$\tilde{b} = b - \frac{1}{2}\sum_{j=1}^{q} \sigma'_j \sigma_j \quad \text{and} \quad \Delta_j w(u,v) := w_j(v) - w_j(u)\ (j = 1, \ldots, q;\ u, v \in [t_0, T])$$

we have, using method (M1)′,

$$\begin{aligned}
&E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{M}}(t) - \bar{y}^{\mathrm{M}}(t) \|^p \\
&\leqslant K \Bigg\{ E \sup_{t_0 \leqslant t \leqslant T} \left\| \int_{[t]^*_G}^{t} \tilde{b}(y^{\mathrm{M}}([t]_G))\, \mathrm{d}s \right\|^p + \sum_{j=1}^{q} E \sup_{t_0 \leqslant t \leqslant T} \left\| \int_{[t]^*_G}^{t} \sigma_j(y^{\mathrm{M}}([t]_G))\, \mathrm{d}s \right\|^p \\
&\quad + \sum_{i,j=1}^{q} E \sup_{t_0 \leqslant t \leqslant T} \| (\sigma'_j \sigma_j)(y^{\mathrm{M}}([t]_G)) [\Delta_i w([t]_G, t) \Delta_j w([t]_G, t) \\
&\qquad - \Delta_i w([t]_G, [t]^*_G) \Delta_j w([t]_G, [t]^*_G)] \|^p \Bigg\} \\
&\leqslant K \left(\frac{h}{m(h)} \right)^{p/2} \left(1 + \ln\left(\frac{m(h)}{h} \right) \right)^{p/2} \\
&\quad + K \sum_{i,j=1}^{q} E \sup_{t_0 \leqslant t \leqslant T} | \Delta_i w([t]_G, t) \Delta_j w([t]_G, t) - \Delta_i w([t]_G, [t]^*_G) \Delta_j w([t]_G, [t]^*_G) |^p,
\end{aligned} \tag{26}$$

analogously to (22)–(24), but having used the inequalities

$$\| \tilde{b}(x) \| \leqslant K(1 + \|x\|)\ (x \in \mathbb{R}^d) \quad \text{and} \quad E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{M}}(t) \|^p \leqslant K.$$

By the Cauchy–Schwarz inequality, by the relations

$$\sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]^*_G, t)|^{2p} = \max_{\substack{0 \leqslant k \leqslant n-1 \\ 0 \leqslant i \leqslant m_k - 1}} \ \sup_{u^k_i \leqslant t \leqslant u^k_{i+1}} |\Delta w_j(u^k_i, t)|^{2p},$$

$$\sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]_G, [t]^*_G)|^{2p} \leqslant \sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]_G, t)|^{2p} = \max_{0 \leqslant k \leqslant n-1} \ \sup_{t_k \leqslant t \leqslant t_{k+1}} |\Delta w_j(t_k, t)|^{2p},$$

and by Lemma 3.3 and (G3) we obtain

$$\begin{aligned}
&E \sup_{t_0 \leqslant t \leqslant T} | \Delta w_i([t]_G, t) \Delta w_j([t]_G, t) - \Delta w_i([t]_G, [t]^*_G) \Delta w_j([t]_G, [t]^*_G) |^p \\
&\leqslant K \Bigg\{ E \sup_{t_0 \leqslant t \leqslant T} | \Delta w_i([t]_G, t) [\Delta w_j([t]_G, t) - \Delta w_j([t]_G, [t]^*_G)] |^p \\
&\quad + E \sup_{t_0 \leqslant t \leqslant T} | [\Delta w_i([t]_G, t) - \Delta w_i([t]_G, [t]^*_G)] \Delta w_j([t]_G, [t]^*_G) |^p \Bigg\} \\
&\leqslant K \Bigg\{ E \left[\sup_{t_0 \leqslant t \leqslant T} | \Delta w_i([t]_G, t) |^p \sup_{t_0 \leqslant t \leqslant T} | \Delta w_j([t]^*_G, t) |^p \right]
\end{aligned}$$

$$
\begin{aligned}
&+ E\left[\sup_{t_0 \leqslant t \leqslant T} |\Delta w_i([t]^*_G, t)|^p \sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]_G, [t]^*_G)|^p\right]\Bigg\} \\
&\leqslant K \left\{\left(E \sup_{t_0 \leqslant t \leqslant T} |\Delta w_i([t]_G, t)|^{2p}\right)^{1/2} \left(E \sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]^*_G, t)|^{2p}\right)^{1/2}\right. \\
&\qquad \left. + \left(E \sup_{t_0 \leqslant t \leqslant T} |\Delta w_i([t]^*_G, t)|^{2p}\right)^{1/2} \left(E \sup_{t_0 \leqslant t \leqslant T} |\Delta w_j([t]_G, [t]^*_G)|^{2p}\right)^{1/2}\right\} \\
&\leqslant K \left\{ \max_{0 \leqslant k \leqslant n-1} h_k^{p/2} \max_{0 \leqslant k \leqslant n-1} \left(\frac{h_k}{m_k}\right)^{p/2} + \max_{0 \leqslant k \leqslant n-1} \left(\frac{h_k}{m_k}\right)^{p/2} \max_{0 \leqslant k \leqslant n-1} h_k^{p/2} \right\} \\
&\quad \times (1 + \ln n + \ln m(h))^{p/2} (1 + \ln n)^{p/2} \\
&\leqslant K \cdot h^{p/2} \left(\frac{h}{m(h)}\right)^{p/2} (1 + \ln n + \ln m(h))^{p/2} (1 + \ln n)^{p/2} \\
&\leqslant K \cdot \left(\frac{h}{m(h)}\right)^{p/2} \left(1 + \ln\left(\frac{m(h)}{h}\right)\right)^{p/2},
\end{aligned} \tag{27}
$$

where the last step is based on (G1) and the boundedness of $h(1 + \ln n) \leqslant h(1 + \ln(\Lambda/h))$. Now, (26) and (27) yield

$$
E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{M}}(t) - \bar{y}^{\mathrm{M}}(t) \|^p \leqslant K \cdot \left(\frac{h}{m(h)}\right)^{p/2} \left(1 + \ln\left(\frac{m(h)}{h}\right)\right)^{p/2}. \tag{28}
$$

Analogously to (25), it follows that

$$
E \sup_{t_0 \leqslant t \leqslant T} \| \bar{y}^{\mathrm{M}}(t) - \tilde{y}^{\mathrm{M}}(t) \|^p \leqslant E \sup_{t_0 \leqslant t \leqslant T} \| y^{\mathrm{M}}(t) - \bar{y}^{\mathrm{M}}(t) \|^p
$$

which, together with (28), gives us the estimate (b). □

4. Chance discretization

In this last discretization step the Wiener process increments shall be replaced by i.i.d. r.v.'s with a given distribution μ on $\mathbb{R}$. But the corresponding results hold only in the weak sense, i.e., the Wiener process (and its increments between the points of G) and i.i.d. r.v.'s ξ^k_{ji} can be defined on a common probability space such that the estimates hold. This applies to Theorem 4.4 being the main result in this section and Theorem 4.1 which was proved in [12, 13] and provides the essential tool for the proof of Theorem 4.4.

Theorem 4.1 (Komlós, Major, Tusnády). *Let $\mu \in \mathscr{P}(\mathbb{R})$ have the following properties:*

$$
\int_{-\infty}^{\infty} x \, \mathrm{d}\mu(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 \, \mathrm{d}\mu(x) = 1 \quad \textit{and} \quad \int_{-\infty}^{\infty} \mathrm{e}^{tx} \, \mathrm{d}\mu(x) < \infty \tag{29}
$$

for all t with $\| t \| \leqslant \tau$, $\tau > 0$.

Then there exist positive constants C, A, λ, only depending on μ, and for each natural number $s > 0$ two s-tuples $(x_1, \ldots, x_s)$ and $(y_1, \ldots, y_s)$, each consisting of i.i.d. real-valued r.v.'s with $D(x_1) = \mu$ and y_1 being standard-normally distributed, such that for each $a > 0$

$$P\left(\max_{1 \leqslant k \leqslant s} \left| \sum_{i=1}^{k} (x_i - y_i) \right| > C \ln s + a\right) < A\mathrm{e}^{-\lambda a}.$$

For translating this estimate into an estimate with the distance used in the previous sections, we need the following lemma.

Lemma 4.2. *Assume, there exist constants $C, A, \lambda > 0$ with $\lambda C \geqslant 1$ and for any two natural numbers $r, s \geqslant 1$ an r-tuple $(\delta_{1,s}, \ldots, \delta_{r,s})$ of i.i.d. positive real-valued r.v.'s satisfying*

$$P(\delta_{1,s} > C \ln s + a) < A\mathrm{e}^{-\lambda a} \quad \text{for all } a > 0. \tag{30}$$

Then for each $p \in [2, \infty)$ there exists a constant $M_p > 0$ such that for all natural $r, s \geqslant 1$

$$E \max_{1 \leqslant i \leqslant r} \delta_{i,s}^p \leqslant M_p (1 + \ln r + \ln s)^p.$$

Proof. With (30) and $z := (C \ln s + a)^p$ we get for $i = 1, \ldots, r$ and $a > 0$

$$P(\delta_{i,s}^p > z) < A\mathrm{e}^{-\lambda a} = A\mathrm{e}^{-\lambda z^{1/p} + \lambda C \ln s} = A s^{\lambda C} \mathrm{e}^{-\lambda z^{1/p}}.$$

This leads to

$$\begin{aligned}
E \max_{1 \leqslant i \leqslant r} \delta_{i,s}^p &= \int_0^\infty P\left(\max_{1 \leqslant i \leqslant r} \delta_{i,s}^p > z\right) \mathrm{d}z \\
&= \int_0^{C^p(1+\ln r+\ln s)^p} P\left(\max_{1 \leqslant i \leqslant r} \delta_{i,s}^p > z\right) \mathrm{d}z + \int_{C^p(1+\ln r+\ln s)^p}^\infty P\left(\max_{1 \leqslant i \leqslant r} \delta_{i,s}^p > z\right) \mathrm{d}z \\
&\leqslant C^p(1 + \ln r + \ln s)^p + r \int_{C^p(1+\ln r+\ln s)^p}^\infty P(\delta_{1,s}^p > z)\, \mathrm{d}z \\
&\leqslant C^p(1 + \ln r + \ln s)^p + A s^{\lambda C} r \int_{C^p(1+\ln r+\ln s)^p}^\infty \mathrm{e}^{-\lambda z^{1/p}} \mathrm{d}z,
\end{aligned}$$

and after substituting $z = C^p(1 + \ln r + \ln s)^p v^p$ we get

$$E \max_{1 \leqslant i \leqslant r} \delta_{i,s}^p \leqslant C^p(1 + \ln r + \ln s)^p \left[1 + A s^{\lambda C} r p \int_1^\infty \mathrm{e}^{-\lambda C(1+\ln r+\ln s)v} v^{p-1} \,\mathrm{d}v \right].$$

It remains to show that the term in the brackets is uniformly bounded for all $r, s \geqslant 1$. Indeed,

$$\begin{aligned}
s^{\lambda C} r \int_1^\infty \mathrm{e}^{-\lambda C(1+\ln r+\ln s)v} v^{p-1} \,\mathrm{d}v &\leqslant s^{\lambda C} r \int_1^\infty \mathrm{e}^{-\lambda C v} s^{-\lambda C v} r^{-\lambda C v} v^{p-1} \,\mathrm{d}v \\
&\leqslant s^{\lambda C} r \int_1^\infty \mathrm{e}^{-v} s^{-\lambda C} r^{-\lambda C} v^{p-1} \,\mathrm{d}v \\
&\leqslant \int_1^\infty \mathrm{e}^{-v} v^{p-1} \,\mathrm{d}v \leqslant \Gamma(p). \qquad \square
\end{aligned}$$

Moreover, we shall make use of the following general result, a proof of which can be found, e.g., in [23, Lemma 2, Theorem 5].

Lemma 4.3. *Let S_1, S_2 and S_3 be polish spaces (i.e., topological spaces which are metrizable with a complete separable metric), and let $P_{12}: S_1 \times S_2 \times S_3 \to S_1 \times S_2$, $P_{23}: S_1 \times S_2 \times S_3 \to S_2 \times S_3$, $P_2^{12}: S_1 \times S_2 \to S_2$ and $P_2^{23}: S_2 \times S_3 \to S_2$ denote the projections defined by dropping one component. Then for any two measures $\nu_{12} \in \mathscr{P}(S_1 \times S_2)$ and $\nu_{23} \in \mathscr{P}(S_2 \times S_3)$ with $\nu_{12} \circ (P_2^{12})^{-1} = \nu_{23} \circ (P_2^{23})^{-1}$, i.e., with identical marginal distributions on S_2, there exists a measure $\nu_{123} \in \mathscr{P}(S_1 \times S_2 \times S_3)$ with $\nu_{123} \circ (P_{12})^{-1} = \nu_{12}$ and $\nu_{123} \circ (P_{23})^{-1} = \nu_{23}$.*

Now we can prove the estimates for the chance discretization step.

Theorem 4.4. *Let $p \in [2, \infty)$ and $\mu \in \mathscr{P}(\mathbb{R})$ have the properties* (29). *Then we can define a q-dimensional standard Wiener process $(w(t))_{t \in [t_0, T]}$ and a set of i.i.d. r.v.'s $\{\xi_{ji}^k\colon j = 1, \ldots, q;\ i = 1, \ldots, m_k;\ k = 0, \ldots, n-1\}$ with distribution $D(\xi_{11}^0) = \mu$ on a common probability space, such that for the methods* (E2), (E3), (M2) *and* (M3) *constructed with them we have*:

(a) *If* (V1) *and* (V2) *hold then*

$$E \sup_{t_0 \leqslant t \leqslant T} \| \tilde{y}^{\mathrm{E}}(t) - z^{\mathrm{E}}(t) \|^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p.$$

(b) *If* (V1)–(V4) *hold then*

$$E \sup_{t_0 \leqslant t \leqslant T} \| \tilde{y}^{\mathrm{M}}(t) - z^{\mathrm{M}}(t) \|^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p.$$

Proof. First we consider a standard Wiener process $(w(t))_{t \in [t_0, T]} = (w_1(t), \ldots, w_q(t))_{t \in [t_0, T]}$ and denote its increments by

$$\Delta_i^k w_j := w_j(u_i^k) - w_j(u_{i-1}^k)$$

for $j = 1, \ldots, q$; $i = 1, \ldots, m_k$; $k = 0, \ldots, n-1$. Further, we denote the vectors of the normed increments within one interval of the coarse grid by

$$\Delta^k w_j^* := \sqrt{\frac{m_k}{h_k}} (\Delta_1^k w_j, \ldots, \Delta_{m_k}^k w_j)^{\mathrm{T}} \in \mathbb{R}^{m_k}$$

for $j = 1, \ldots, q$; $k = 0, \ldots, n-1$ and gather them in the vector

$$\Delta w^* := ((\Delta^0 w_1^*)^{\mathrm{T}}, (\Delta^1 w_1^*)^{\mathrm{T}}, \ldots, (\Delta^{n-1} w_1^*)^{\mathrm{T}}, \ldots, (\Delta^0 w_q^*)^{\mathrm{T}}, (\Delta^1 w_q^*)^{\mathrm{T}}, \ldots, (\Delta^{n-1} w_q^*)^{\mathrm{T}})^{\mathrm{T}}.$$

Hence, for the common distribution of w and Δw^* we have $\nu_{12} := D(w, \Delta w^*) \in \mathscr{P}(S_1 \times S_2)$ where $S_1 = C([t_0, T]; \mathbb{R}^d)$ and $S_2 = \mathbb{R}^{q \sum_{k=0}^{n-1} m_k}$. Let $\hat{m} := \max_{k=0,\ldots,n-1} m_k$ and consider a measure $\mu \in \mathscr{P}(\mathbb{R})$ with (29). Then, by Theorem 4.1 and Lemma 4.2, we have the existence of i.i.d. r.v.'s $x_1, \ldots, x_{\hat{m}}$ with distribution $D(x_1) = \mu$ and of independent standard-normally distributed r.v.'s

$y_1, \dots, y_{\hat{m}}$, such that for suitable constants $C, A, \lambda > 0$ which only depend on μ it holds that

$$P\left(\max_{1 \leqslant i \leqslant \hat{m}} \left| \sum_{s=1}^{i} x_s - \sum_{s=1}^{i} y_s \right| \geqslant C \ln \hat{m} + a \right) < A e^{-\lambda a} \tag{31}$$

for all $a > 0$. We define for $k = 0, \dots, n-1$ the joint distribution

$$\eta_k := D((y_1, \dots, y_{m_k}, x_1, \dots, x_{m_k})^{\mathrm{T}}) \in \mathscr{P}(\mathbb{R}^{m_k} \times \mathbb{R}^{m_k})$$

and the q-fold product measure of η_k:

$$\eta_k^q := \eta_k \otimes \cdots \otimes \eta_k \in \mathscr{P}(\mathbb{R}^{q m_k} \times \mathbb{R}^{q m_k}).$$

Finally, we define

$$\nu_{23} := \eta_0^q \otimes \eta_1^q \otimes \eta_2^q \otimes \cdots \otimes \eta_{n-1}^q \in \mathscr{P}(S_2 \times S_3),$$

where $S_2 = S_3 = \mathbb{R}^{q \sum_{k=0}^{n-1} m_k}$, its projection to $\mathscr{P}(S_2)$ being $\eta_2 = \otimes_{k=0}^{n-1} D((y_1, \dots, y_{m_k})^{\mathrm{T}}) \in \mathscr{P}(S_2)$. Then, by Lemma 4.3, there is a measure $\nu_{123} \in \mathscr{P}(S_1 \times S_2 \times S_3)$ with projections $\nu_{12} \in \mathscr{P}(S_1 \times S_2)$ and $\nu_{23} \in \mathscr{P}(S_2 \times S_3)$. We choose a random variable

$$\gamma = (\tilde{w}, \Delta \tilde{w}^*, \xi) : (\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P}) \to S_1 \times S_2 \times S_3 \quad \text{with } D(\gamma) = \nu_{123},$$

$\tilde{w}$ and $\Delta \tilde{w}^*$ being a standard Wiener process and its vector of normed increments on the grid G, respectively, and, thus, having the same joint distribution as $(w, \Delta w^*)$. For this reason we can write, without limitation of generality, $(w, \Delta w^*)$ instead of $(\tilde{w}, \Delta \tilde{w}^*)$. ξ has the following shape:

$$\xi := ((\xi_1^0)^{\mathrm{T}}, (\xi_1^1)^{\mathrm{T}}, \dots, (\xi_1^{n-1})^{\mathrm{T}}, \dots, (\xi_q^0)^{\mathrm{T}}, (\xi_q^1)^{\mathrm{T}}, \dots, (\xi_q^{n-1})^{\mathrm{T}})^{\mathrm{T}}$$

with $\xi_j^k := (\xi_{j1}^k, \dots, \xi_{jm_k}^k)^{\mathrm{T}}$, $k = 0, \dots, n-1$; $j = 1, \dots, q$. This means, ξ consists of $q \sum_{k=0}^{n-1} m_k$ i.i.d. components, each with distribution μ. For a while, let us extend the vectors

$$\Delta w^* = (\Delta_i^k w_j)_{j=1,\dots,q;\, k=0,\dots,n-1;\, i=1,\dots,m_k} \quad \text{and} \quad \xi = (\xi_{ji}^k)_{j=1,\dots,q;\, k=0,\dots,n-1;\, i=1,\dots,m_k}$$

to

$$(\Delta_i^k w_j)_{j=1,\dots,q;\, k=0,\dots,n-1;\, i=1,\dots,\hat{m}} \quad \text{and} \quad (\xi_{ji}^k)_{j=1,\dots,q;\, k=0,\dots,n-1;\, i=1,\dots,\hat{m}},$$

which also consist of i.i.d. components. Then, since

$$D((x_1, \dots, x_{m_k}, y_1, \dots, y_{m_k})^{\mathrm{T}}) = D\left(\left(\xi_{j1}^k, \dots, \xi_{jm_k}^k, \sqrt{\frac{m_k}{h_k}} \Delta_1^k w_j, \dots, \sqrt{\frac{m_k}{h_k}} \Delta_{m_k}^k w_j\right)^{\mathrm{T}}\right),$$

(31) yields for all $k = 0, \dots, n-1$ and $j = 1, \dots, q$ that

$$P\left(\max_{1 \leqslant i \leqslant \hat{m}} \left| \sum_{s=1}^{i} \xi_{js}^k - \sum_{s=1}^{i} \sqrt{\frac{m_k}{h_k}} \Delta_s^k w_j \right| > C \ln \hat{m} + a \right) < A e^{-\lambda a} \quad \text{for all } a > 0,$$

and now we can apply Lemma 4.2 and get with $m_k \leqslant \hat{m} \leqslant m(h)^{\alpha}$ (see (G2)) the two estimates

$$E \max_{0 \leqslant i \leqslant m_k} \left| \sum_{s=1}^{i} \xi_{js}^k - \sum_{s=1}^{i} \sqrt{\frac{m_k}{h_k}} \Delta_s^k w_j \right|^p \leqslant K(1 + \ln m(h))^p \tag{32}$$

for $k = 0, \dots, n-1$, $j = 1, \dots, q$ and

$$E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \left| \sum_{s=1}^{i} \xi_{js}^{k} - \sum_{s=1}^{i} \sqrt{\frac{m_k}{h_k}} \Delta_s^k w_j \right|^p \leqslant K(1 + \ln n + \ln m(h))^p \tag{33}$$

for $j = 1, \dots, q$. Moreover, we have by construction that for $k = 1, \dots, n-1$

ξ_{js}^r und $\Delta_s^r w_j$, $j = 1, \dots, q$; $s = 1, \dots, m_r$; $r = k, \dots, n-1$,

are independent of the σ-algebra $\mathscr{A}_k$ generated by

$$\{\xi_{js}^r, \Delta_s^r w_j \colon j = 1, \dots, q;\ s = 1, \dots, m_r;\ r = 0, \dots, k-1\}. \tag{34}$$

Now we consider w and ξ, as constructed above, as well as the approximation methods (E2), (E3), (M2) and (M3) defined on the basis of w and ξ. According to the definitions, for the estimates (a) and (b) only the values of the approximate solutions in the grid points of G have to be taken into account:

(a) First we consider the approximate solutions (E2) and (E3) only in the grid points t_k of the coarse subgrid of G. Then, with the notation

$$\Delta_k w_j := w_j(t_{k+1}) - w_j(t_k), \quad j = 1, \dots, q;\ k = 0, \dots, n-1,$$

the definitions of (E2) and (E3) yield, with Lemma 2.1(b), for $k = 0, \dots, n$

$$\begin{aligned} E \max_{0 \leqslant f \leqslant k} \| \tilde{y}^{\mathrm{E}}(t_f) - z^{\mathrm{E}}(t_f) \|^p \leqslant K \Bigg\{ & E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} h_r [b(\tilde{y}^{\mathrm{E}}(t_r)) - b(z^{\mathrm{E}}(t_r))] \right\|^p \\ & + \sum_{j=1}^{q} E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} \left[\sigma_j(\tilde{y}^{\mathrm{E}}(t_r)) \Delta_r w_j - \sigma(z^{\mathrm{E}}(t_r)) \sqrt{\frac{h_r}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r \right] \right\|^p \Bigg\} \\ =: K \Bigg\{ & D_1^{\mathrm{E}}(k) + \sum_{j=1}^{q} D_{2j}^{\mathrm{E}}(k) \Bigg\}. \end{aligned} \tag{35}$$

Now, with Lemma 2.1(b), (V2) and (G1), it follows that for $k = 0, \dots, n$,

$$\begin{aligned} D_1^{\mathrm{E}}(k) &\leqslant k^{p-1} L^p E \max_{0 \leqslant f \leqslant k} \sum_{r=0}^{f-1} h_r^p \| \tilde{y}^{\mathrm{E}}(t_r) - z^{\mathrm{E}}(t_r) \|^p \leqslant K (nh)^p \frac{1}{n} E \sum_{r=0}^{k-1} \| \tilde{y}^{\mathrm{E}}(t_r) - z^{\mathrm{E}}(t_r) \|^p \\ &\leqslant K \cdot \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{E}}(t_s) - z^{\mathrm{E}}(t_s) \|^p \end{aligned} \tag{36}$$

and, for $j = 1, \dots, q$,

$$\begin{aligned} D_{2j}^{\mathrm{E}}(k) \leqslant K \Bigg\{ & E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} [\sigma_j(\tilde{y}^{\mathrm{E}}(t_r)) - \sigma_j(z^{\mathrm{E}}(t_r))] \Delta_r w_j \right\|^p \\ & + E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} \sigma_j(z^{\mathrm{E}}(t_r)) \left[\Delta_r w_j - \sqrt{\frac{h_r}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r \right] \right\|^p \Bigg\} \\ =: K \{ & D_{2j1}^{\mathrm{E}}(k) + D_{2j2}^{\mathrm{E}}(k) \}. \end{aligned} \tag{37}$$

Because of (34),

$$M_{j1}(f) := \sum_{r=0}^{f-1} [\sigma_j(\tilde{y}^{\mathrm{E}}(t_r)) - \sigma_j(z^{\mathrm{E}}(t_r))] \Delta_r w_j,$$

$$M_{j2}(f) := \sum_{r=0}^{f-1} \sigma_j(z^{\mathrm{E}}(t_r)) \left[\Delta_r w_j - \sqrt{\frac{h_r}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r \right], \quad f = 0, \dots, n,$$

are d-dimensional martingales w.r.t. $(\mathscr{A}_f)_{f=0,\dots,n}$, and that is why, using Corollary 2.4(b), they can be estimated in the following way for all $j = 1, \dots, q$ and $k = 0, \dots, n$:

$$D_{2j1}^{\mathrm{E}}(k) \leqslant K(dk)^{p/2-1} \sum_{r=0}^{k-1} E\{\| \sigma_j(\tilde{y}^{\mathrm{E}}(t_r)) - \sigma_j(z^{\mathrm{E}}(t_r)) \|^p |\Delta_r w_j|^p\}.$$

Since both factors in the braces are independent (because of (34)), with (V2) and (G1) it follows that

$$\begin{aligned} D_{2j1}^{\mathrm{E}}(k) &\leqslant K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \{E\| \sigma_j(\tilde{y}^{\mathrm{E}}(t_r)) - \sigma_j(z^{\mathrm{E}}(t_r)) \|^p E|\Delta_r w_j|^p\} \\ &\leqslant K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \left\{ h_r^{p/2} E \left(\frac{1}{\sqrt{h_r}} |\Delta_r w_j| \right)^p E\| \tilde{y}^{\mathrm{E}}(t_r) - z^{\mathrm{E}}(t_r) \|^p \right\} \\ &\leqslant K \cdot n^{p/2-1} h^{p/2} \sum_{r=0}^{k-1} E\| \tilde{y}^{\mathrm{E}}(t_r) - z^{\mathrm{E}}(t_r) \|^p \\ &\leqslant K \cdot \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{E}}(t_s) - z^{\mathrm{E}}(t_s) \|^p. \end{aligned} \tag{38}$$

Here we used that, since all $(1/\sqrt{h_r})\Delta_r w_j$ are standard-normally distributed, all $E|(1/\sqrt{h_r})\Delta_r w_j|^p$ are equal to the same constant only depending on p.

For the other summand Corollary 2.4(b), (V1), (32) and (G1) and (G3) yield

$$\begin{aligned} D_{2j2}^{\mathrm{E}}(k) &\leqslant K \cdot (dk)^{p/2-1} \sum_{r=0}^{k-1} E \left\{ \| \sigma_j(z^{\mathrm{E}}(t_r)) \|^p \left| \Delta_r w_j - \sqrt{\frac{h_r}{m_r}} \sum_{s=1}^{m_r} \xi_{js}^r \right|^p \right\} \\ &\leqslant K \cdot n^{p/2-1} \sum_{r=0}^{k-1} \left(\sqrt{\frac{h_r}{m_r}} \right)^p E \left| \sqrt{\frac{m_r}{h_r}} \Delta_r w_j - \sum_{s=1}^{m_r} \xi_{js}^r \right|^p \\ &\leqslant K \cdot n^{p/2-1} \left(\frac{h}{m(h)} \right)^{p/2} k (1 + \ln m(h))^p \\ &\leqslant K \left(\frac{nh}{m(h)} \right)^{p/2} (1 + \ln m(h))^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p. \end{aligned} \tag{39}$$

Now, considering (35)–(39), we get for all $k = 1, \dots, n$ that

$$E \max_{0 \leqslant f \leqslant k} \| \tilde{y}^{\mathrm{E}}(t_f) - z^{\mathrm{E}}(t_f) \|^p \leqslant K \left\{ \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{E}}(t_s) - z^{\mathrm{E}}(t_s) \|^p + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \right\}$$

and with Lemma 2.5(b) we have

$$E \max_{0 \leqslant f \leqslant n} \| \tilde{y}^{\mathrm{E}}(t_f) - z^{\mathrm{E}}(t_f) \|^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p. \tag{40}$$

In the next step we extend this estimate to the intermediate grid points u_i^k, and we use the notation $\Delta_{i,0}^k w_j := w_j(u_i^k) - w_j(u_0^k)$, $j = 1, \ldots, q$; $k = 0, \ldots, n-1$; $i = 1, \ldots, m_k$:

$$\begin{aligned}
& E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \| \tilde{y}^{\mathrm{E}}(u_i^k) - z^{\mathrm{E}}(u_i^k) \|^p \\
& \leqslant K \left\{ E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{E}}(t_k) - z^{\mathrm{E}}(t_k) \|^p + E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \left\| h_k \cdot \frac{i}{m_k} (b(\tilde{y}^{\mathrm{E}}(t_k)) - b(z^{\mathrm{E}}(t_k))) \right\|^p \right. \\
& \qquad \left. + \sum_{j=1}^q E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \left\| \sigma_j(\tilde{y}^{\mathrm{E}}(t_k)) \Delta_{i,0}^k w_j - \sigma_j(z^{\mathrm{E}}(t_k)) \sqrt{\frac{h_k}{m_k}} \sum_{s=1}^i \xi_{js}^k \right\|^p \right\} \\
& =: K \left\{ E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{E}}(t_k) - z^{\mathrm{E}}(t_k) \|^p + D_4^{\mathrm{E}} + \sum_{j=1}^q D_{5j}^{\mathrm{E}} \right\}.
\end{aligned} \tag{41}$$

Here, (V2) implies

$$D_4^{\mathrm{E}} \leqslant K \cdot h^p E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{E}}(t_k) - z^{\mathrm{E}}(t_k) \|^p. \tag{42}$$

On the other hand, we have for all $j = 1, \ldots, q$ that

$$\begin{aligned}
D_{5j}^{\mathrm{E}} & \leqslant K \left\{ E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \| [\sigma_j(\tilde{y}^{\mathrm{E}}(t_k)) - \sigma_j(z^{\mathrm{E}}(t_k))] \Delta_{i,0}^k w_j \|^p \right. \\
& \qquad \left. + E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \left\| \sigma_j(z^{\mathrm{E}}(t_k)) \left[\Delta_{i,0}^k w_j - \sqrt{\frac{h_k}{m_k}} \sum_{s=1}^i \xi_{js}^k \right] \right\|^p \right\} \\
& =: K \{ D_{5j1}^{\mathrm{E}} + D_{5j2}^{\mathrm{E}} \}.
\end{aligned} \tag{43}$$

Further, using (V2), the Cauchy–Schwarz inequality, (40), Lemma 3.3 and (G1), we get

$$\begin{aligned}
D_{5j1}^{\mathrm{E}} & \leqslant E \left\{ \left(\max_{0 \leqslant k \leqslant n-1} \| \sigma_j(\tilde{y}^{\mathrm{E}}(t_k)) - \sigma_j(z^{\mathrm{E}}(t_k)) \|^p \right) \left(\max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} | \Delta_{i,0}^k w_j |^p \right) \right\} \\
& \leqslant K \left(E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{E}}(t_k) - z^{\mathrm{E}}(t_k) \|^{2p} \right)^{1/2} \left(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} | \Delta_{i,0}^k w_j |^{2p} \right)^{1/2} \\
& \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \left(E \max_{0 \leqslant k \leqslant n-1} \sup_{t_k \leqslant t \leqslant t_{k+1}} | w_j(t) - w_j(t_k) |^{2p} \right)^{1/2} \\
& \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p h^{p/2} (1 + \ln n)^{p/2} \\
& \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p,
\end{aligned} \tag{44}$$

since $h(1+\ln n) \leqslant K \cdot h(1+\ln(\Lambda/h)) \leqslant K$ for $h \in (0, T-t_0]$. This estimate was done so roughly since, for the final result, here a better estimate than in (40) does not pay. This consideration applies also to the following estimate: With (V1), (G3), (33) and (G1) it follows that

$$D_{5j2}^{\mathrm{E}} \leqslant K \cdot E\left\{\max_{0\leqslant k\leqslant n-1}\max_{0\leqslant i\leqslant m_k}\left(\frac{h_k}{m_k}\right)^{p/2}\left\|\sqrt{\frac{m_k}{h_k}}\,\Delta_{i,0}^k w_j - \sum_{s=1}^{i}\xi_{js}^k\right\|^p\right\}$$

$$\leqslant K\left(\frac{h}{m(h)}\right)^{p/2}(1+\ln n+\ln m(h))^p \leqslant K\left(\frac{1+\ln n+\ln m(h)}{\sqrt{n\cdot m(h)}}\right)^p$$

$$\leqslant K\left(\frac{1+\ln m(h)}{\sqrt{m(h)}}\right)^p. \tag{45}$$

The last step is implied by

$$\frac{1+\ln\gamma+\ln\delta}{\sqrt{\gamma\delta}} \leqslant \frac{1+\ln\gamma+\ln\delta+\ln\gamma\ln\delta}{\sqrt{\gamma\delta}} = \left(\frac{1+\ln\gamma}{\sqrt{\gamma}}\right)\left(\frac{1+\ln\delta}{\sqrt{\delta}}\right)$$

$$\leqslant \frac{2}{\sqrt{\mathrm{e}}}\left(\frac{1+\ln\delta}{\sqrt{\delta}}\right) \quad \text{for all real } \gamma,\ \delta \geqslant 1. \tag{46}$$

Now it follows from (41)–(45) that

$$E\max_{0\leqslant k\leqslant n-1}\max_{0\leqslant i\leqslant m_k}\|\tilde{y}^{\mathrm{E}}(u_i^k) - z^{\mathrm{E}}(u_i^k)\|^p \leqslant K\left\{E\max_{0\leqslant k\leqslant n-1}\|\tilde{y}^{\mathrm{E}}(t_k) - z^{\mathrm{E}}(t_k)\|^p + \left(\frac{1+\ln m(h)}{\sqrt{m(h)}}\right)^p\right\},$$

and with (40) we get assertion (a).

(b) The formulae (M2) and (M3) yield, with Lemma 2.1(b), for all $k = 0, \dots, n$,

$$E\max_{0\leqslant f\leqslant k}\|\tilde{y}^{\mathrm{M}}(t_f) - z^{\mathrm{M}}(t_f)\|^p$$

$$\leqslant K\left\{E\max_{0\leqslant f\leqslant k}\|\tilde{b}(\tilde{y}^{\mathrm{M}}(t_f)) - \tilde{b}(z^{\mathrm{M}}(t_f))\|^p\right.$$

$$+\sum_{j=1}^{q}E\max_{0\leqslant f\leqslant k}\left\|\sum_{r=0}^{f-1}\left[\sigma_j(\tilde{y}^{\mathrm{M}}(t_r))\Delta_r w_j - \sigma_j(z^{\mathrm{M}}(t_r))\sqrt{\frac{h_k}{m_k}}\sum_{s=1}^{m_r}\xi_{js}^r\right]\right\|^p$$

$$\left.+\frac{1}{2^p}\sum_{j,g=1}^{q}E\max_{0\leqslant f\leqslant k}\left\|\sum_{r=0}^{f-1}\left[(\sigma_j'\sigma_g)(\tilde{y}^{\mathrm{M}}(t_r))\Delta_r w_j\Delta_r w_g - (\sigma_j'\sigma_g)(z^{\mathrm{M}}(t_r))\frac{h_r}{m_r}\sum_{s=1}^{m_r}\xi_{js}^r\sum_{s=1}^{m_r}\xi_{gs}^r\right]\right\|^p\right\}$$

$$=: K\left\{D_1^{\mathrm{M}}(k) + \sum_{j=1}^{q}D_{2j}^{\mathrm{M}}(k) + \frac{1}{2^p}\sum_{j,g=1}^{q}D_{3jg}^{\mathrm{M}}(k)\right\}, \tag{47}$$

where $\tilde{b} := b - \frac{1}{2}\sum_{j=1}^{q} \sigma'_j \sigma_j$ is Lipschitz continuous because of (V1)–(V3). Now we get, analogously to the estimates of $D_1^{\mathrm{E}}(k)$ and $D_{2j}^{\mathrm{E}}(k)$, for all $k = 1, \ldots, n$ and $j = 1, \ldots, q$,

$$D_1^{\mathrm{M}}(k) \leqslant K \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{M}}(t_s) - z^{\mathrm{M}}(t_s) \|^p, \tag{48}$$

$$D_{2j}^{\mathrm{M}}(k) \leqslant K \left\{ \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{M}}(t_s) - z^{\mathrm{M}}(t_s) \|^p + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \right\}. \tag{49}$$

The third kind of summands can be estimated for $k = 1, \ldots, n$ and $j, g = 1, \ldots, q$ by

$$\begin{aligned} D_{3jg}^{\mathrm{M}}(k) &\leqslant K \left\{ E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} [(\sigma'_j \sigma_g)(\tilde{y}^{\mathrm{M}}(t_r)) - (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_r))] \Delta_r w_j \Delta_r w_g \right\|^p \right. \\ &\qquad \left. + E \max_{0 \leqslant f \leqslant k} \left\| \sum_{r=0}^{f-1} (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_r)) \left[\Delta_r w_j \Delta_r w_g - \frac{h_r}{m_r} \sum_{s=1}^{m_r} \xi_{js}^r \sum_{s=1}^{m_r} \xi_{gs}^r \right] \right\|^p \right\} \\ &=: K \{ D_{3jg1}^{\mathrm{M}}(k) + D_{3jg2}^{\mathrm{M}}(k) \}. \end{aligned} \tag{50}$$

Since all $\sigma'_j \sigma_g$ are Lipschitz continuous because of (V1)–(V3) we get with Lemma 2.1(b) and the martingale property of $\tilde{y}^{\mathrm{M}}$ and z^{M} that

$$\begin{aligned} D_{3jg1}^{\mathrm{M}}(k) &\leqslant K \cdot E \max_{0 \leqslant f \leqslant k} \left\{ f^{p-1} \sum_{r=0}^{f-1} \| \tilde{y}^{\mathrm{M}}(t_r) - z^{\mathrm{M}}(t_r) \|^p |\Delta_r w_j|^p |\Delta_r w_g|^p \right\} \\ &\leqslant K \cdot n^{p-1} \sum_{r=0}^{k-1} E \| \tilde{y}^{\mathrm{M}}(t_r) - z^{\mathrm{M}}(t_r) \|^p h_r^p E \left\{ \left| \frac{1}{\sqrt{h_r}} \Delta_r w_j \right|^p \left| \frac{1}{\sqrt{h_r}} \Delta_r w_g \right|^p \right\}. \end{aligned}$$

Taking into account that all $(1/\sqrt{h_r}) \Delta_r w_j$ $(r = 0, \ldots, n-1;\ j = 1, \ldots, q)$ are independent and standard-normally distributed, the last expectation, for $j = g$ as well as for $j \neq g$ (then writing it as the product of the expectations of both factors), turns out to be a constant only depending on p. Thus, with (G1) we get

$$\begin{aligned} D_{3jg1}^{\mathrm{M}}(k) &\leqslant K \cdot n^p h^p \frac{1}{n} \sum_{r=0}^{k-1} E \| \tilde{y}^{\mathrm{M}}(t_r) - z^{\mathrm{M}}(t_r) \|^p \\ &\leqslant K \cdot \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \| \tilde{y}^{\mathrm{M}}(t_s) - z^{\mathrm{M}}(t_s) \|^p. \end{aligned} \tag{51}$$

In the sequel we will use for every $j = 1, \ldots, q$ and $r = 0, \ldots, n-1$ the notation

$$\Delta_r \xi_j := \sqrt{(h_r/m_r)} \textstyle\sum_{s=1}^{m_r} \xi_{js}^r.$$

The boundedness of $\sigma'_j \sigma_g$ and Lemma 2.1(b) yield

$$\begin{aligned} D_{3jg2}^{\mathrm{M}}(k) &\leqslant E \max_{0 \leqslant f \leqslant k} f^{p-1} \sum_{r=0}^{f-1} \| (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_r)) \|^p |\Delta_r w_j \Delta_r w_g - \Delta_r \xi_j \Delta_r \xi_g|^p \\ &\leqslant K \cdot n^{p-1} \sum_{r=0}^{n-1} E |\Delta_r w_j \Delta_r w_g - \Delta_r \xi_j \Delta_r \xi_g|^p. \end{aligned} \tag{52}$$

Using Lemma 2.1(b) and the Cauchy–Schwarz inequality, we get

$$
\begin{aligned}
&E|\Delta_r w_j \Delta_r w_g - \Delta_r \xi_j \Delta_r \xi_g|^p \\
&\quad \leqslant K(E\{|\Delta_r w_j - \Delta_r \xi_j|^p |\Delta_r w_g|^p\} + E\{|\Delta_r \xi_j - \Delta_r w_j|^p |\Delta_r w_g - \Delta_r \xi_g|^p\} \\
&\qquad + E\{|\Delta_r w_j|^p |\Delta_r w_g - \Delta_r \xi_g|^p\}) \\
&\quad \leqslant K \cdot h_r^p \left(\left[E\left|\frac{1}{\sqrt{h_r}}(\Delta_r w_j - \Delta_r \xi_j)\right|^{2p} E\left|\frac{1}{\sqrt{h_r}}\Delta_r w_g\right|^{2p}\right]^{1/2} \right. \\
&\qquad + \left[E\left|\frac{1}{\sqrt{h_r}}(\Delta_r \xi_j - \Delta_r w_j)\right|^{2p} E\left|\frac{1}{\sqrt{h_r}}(\Delta_r w_g - \Delta_r \xi_g)\right|^{2p}\right]^{1/2} \\
&\qquad \left. + \left[E\left|\frac{1}{\sqrt{h_r}}\Delta_r w_j\right|^{2p} E\left|\frac{1}{\sqrt{h_r}}(\Delta_r w_g - \Delta_r \xi_g)\right|^{2p}\right]^{1/2} \right).
\end{aligned}
$$

Since all $(1/\sqrt{h_r})\Delta_r w_j$ are standard-normally distributed, we obtain with (32) and (G3) that

$$E|\Delta_r w_j \Delta_r w_g - \Delta_r \xi_j \Delta_r \xi_g|^p \leqslant K \cdot h_r^p (1 + \ln m(h))^p \left(\frac{1}{\sqrt{m_r}}\right)^p \leqslant K \cdot h_r^{p/2} (1 + \ln m(h))^p \left(\sqrt{\frac{h}{m(h)}}\right)^p,$$

i.e., with (52) and (G1) we have

$$D_{3jg2}^{\mathrm{M}}(k) \leqslant K \cdot n^p h^{p/2} (1 + \ln m(h))^p \left(\sqrt{\frac{h}{m(h)}}\right)^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}}\right)^p. \tag{53}$$

Now, (47)–(51) and (53) yield for every $k = 0, \ldots, n$,

$$E \max_{0 \leqslant f \leqslant k} \|\tilde{y}^{\mathrm{M}}(t_f) - z^{\mathrm{M}}(t_f)\|^p \leqslant K \left\{ \frac{1}{n} \sum_{r=0}^{k-1} E \max_{0 \leqslant s \leqslant r} \|\tilde{y}^{\mathrm{M}}(t_s) - z^{\mathrm{M}}(t_s)\|^p + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}}\right)^p \right\}.$$

Thus, with Lemma 2.5(b) it follows that

$$E \max_{0 \leqslant k \leqslant n} \|\tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k)\|^p \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}}\right)^p. \tag{54}$$

It remains to extend this estimate to the intermediate grid points u_i^k. After introducing the notation

$$\Delta_i^k \xi_j := \sqrt{\frac{h_k}{m_k}} \sum_{s=1}^{i} \xi_{js}^k \quad \text{for } j = 1, \ldots, q;\ i = 1, \ldots, m_k;\ k = 0, \ldots, n-1,$$

we get with the definitions of the methods (M2) and (M3) and with Lemma 2.1(b) that

$$
\begin{aligned}
&E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \|\tilde{y}^{\mathrm{M}}(u_i^k) - z^{\mathrm{M}}(u_i^k)\|^p \\
&\quad \leqslant K \left\{ E \max_{0 \leqslant k \leqslant n-1} \|\tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k)\|^p + E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \left\| h_k \cdot \frac{i}{m_k} [\tilde{b}(\tilde{y}^{\mathrm{M}}(t_k)) - \tilde{b}(z^{\mathrm{M}}(t_k))] \right\|^p \right.
\end{aligned}
$$

$$+ \sum_{j=1}^{q} E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \| \sigma_j(\tilde{y}^{\mathrm{M}}(t_k)) \Delta_{i,0}^k w_j - \sigma_j(z^{\mathrm{M}}(t_k)) \Delta_i^k \xi_j \|^p$$

$$+ \sum_{j,g=1}^{q} E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \| (\sigma'_j \sigma_g)(\tilde{y}^{\mathrm{M}}(t_k)) \Delta_{i,0}^k w_j \Delta_{i,0}^k w_g - (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_k)) \Delta_i^k \xi_j \Delta_i^k \xi_g \|^p \Bigg\}$$

$$=: K \left\{ E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k) \|^p + D_4^{\mathrm{M}} + \sum_{j=1}^{q} D_{5j}^{\mathrm{M}} + \sum_{j,g=1}^{q} D_{6jg}^{\mathrm{M}} \right\}. \tag{55}$$

Because of the Lipschitz continuity of $\tilde{b}$ and using (54), we can obtain, the same way as the estimates for D_4^{E} and D_{5j}^{E} ((42)–(45)), the following estimates:

$$D_4^{\mathrm{M}} \leqslant K \cdot h^p \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p, \tag{56}$$

$$D_{5j}^{\mathrm{M}} \leqslant K \cdot \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \quad \text{for all } j = 1, \ldots, q. \tag{57}$$

For D_{6jg}^{M}, $j, g = 1, \ldots, q$, using the Lipschitz continuity and boundedness of $\sigma'_j \sigma_g$, taking the maxima for each factor and using the Cauchy–Schwarz inequality twice, we get

$$D_{6jg}^{M} \leqslant K \Bigg\{ E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} [\| (\sigma'_j \sigma_g)(\tilde{y}^{\mathrm{M}}(t_k)) - (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_k)) \|^p |\Delta_{i,0}^k w_j|^p |\Delta_{i,0}^k w_g|^p]$$

$$+ E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} [\| (\sigma'_j \sigma_g)(z^{\mathrm{M}}(t_k)) \|^p |\Delta_{i,0}^k w_j \Delta_{i,0}^k w_g - \Delta_i^k \xi_j \Delta_i^k \xi_g|^p] \Bigg\}$$

$$\leqslant K \Bigg\{ \left(E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k) \|^{2p} \right)^{1/2} \left(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j|^{4p} \right)^{1/4}$$

$$\times \left(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_g|^{4p} \right)^{1/4}$$

$$+ E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j \Delta_{i,0}^k w_g - \Delta_i^k \xi_j \Delta_i^k \xi_g|^p \Bigg\}. \tag{58}$$

Lemma 3.3 and (33) yield for $j = 1, \ldots, q$ and $\bar{p} \geqslant 2$ the estimates

$$E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j|^{\bar{p}} \leqslant K(\bar{p}) \cdot h^{\bar{p}/2} (1 + \ln n)^{\bar{p}/2} \leqslant K(\bar{p}) \left(\frac{1 + \ln n}{n} \right)^{\bar{p}/2}, \tag{59}$$

$$E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j - \Delta_i^k \xi_j|^{\bar{p}} \leqslant K(\bar{p}) \left(\frac{h}{m(h)} \right)^{\bar{p}/2} (1 + \ln n + \ln m(h))^{\bar{p}}$$

$$\leqslant K(\bar{p}) \left(\frac{1 + \ln(n \cdot m(h))}{\sqrt{n \cdot m(h)}} \right)^{\bar{p}}. \tag{60}$$

Then, Lemma 2.1(b), the Cauchy–Schwarz inequality, (59) and (60) imply that for $j, g = 1, \ldots, q$ we have

$$\begin{aligned}
&E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j \Delta_{i,0}^k w_g - \Delta_i^k \xi_j \Delta_i^k \xi_g|^p \\
&\leqslant K \Big\{ E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} (|\Delta_{i,0}^k w_j - \Delta_i^k \xi_j|^p |\Delta_{i,0}^k w_g|^p) \\
&\qquad + E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} (|\Delta_i^k \xi_j - \Delta_{i,0}^k w_j|^p |\Delta_{i,0}^k w_g - \Delta_i^k \xi_g|^p) \\
&\qquad + E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} (|\Delta_{i,0}^k w_j|^p |\Delta_{i,0}^k w_g - \Delta_i^k \xi_g|^p) \Big\} \\
&\leqslant K \Big\{ \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j - \Delta_i^k \xi_j|^{2p} \Big)^{1/2} \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_g|^{2p} \Big)^{1/2} \\
&\qquad + \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_i^k \xi_j - \Delta_{i,0}^k w_j|^{2p} \Big)^{1/2} \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_g - \Delta_i^k \xi_g|^{2p} \Big)^{1/2} \\
&\qquad + \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_j|^{2p} \Big)^{1/2} \Big(E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} |\Delta_{i,0}^k w_g - \Delta_i^k \xi_g|^{2p} \Big)^{1/2} \Big\} \\
&\leqslant K \left\{ \left(\frac{1 + \ln(n \cdot m(h))}{\sqrt{n \cdot m(h)}} \right)^p \left(\frac{1 + \ln n}{n} \right)^{p/2} + \left(\frac{1 + \ln(n \cdot m(h))}{\sqrt{n \cdot m(h)}} \right)^{2p} \right\} \\
&\leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p
\end{aligned} \tag{61}$$

for the last step taking into account (46) and that $(1 + \ln x)/x$ and $(1 + \ln x)/\sqrt{x}$ are bounded for all $x \geqslant 1$. Thus, (58), (59) and (61) yield for all $j, g = 1, \ldots, q$,

$$D_{6jg}^{\mathrm{M}} \leqslant K \left\{ \Big(E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k) \|^{2p} \Big)^{1/2} \left(\frac{1 + \ln n}{n} \right)^p + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \right\}. \tag{62}$$

Finally, from (55)–(57), (62) and (54) we can conclude that

$$\begin{aligned}
&E \max_{0 \leqslant k \leqslant n-1} \max_{0 \leqslant i \leqslant m_k} \| \tilde{y}^{\mathrm{M}}(u_i^k) - z^{\mathrm{M}}(u_i^k) \|^p \\
&\leqslant K \Big\{ E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k) \|^p + \Big(E \max_{0 \leqslant k \leqslant n-1} \| \tilde{y}^{\mathrm{M}}(t_k) - z^{\mathrm{M}}(t_k) \|^{2p} \Big)^{1/2} \\
&\qquad + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \Big\} \\
&\leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p,
\end{aligned}$$

and assertion (b) follows. □

5. Tuning of time and chance discretization. Main results

The main results of the preceding three sections (Theorems 2.6, 3.4 and 4.4) yield the following theorem which gives bounds for the L^p-norm of the differences between the exact solution x of (I) and the approximate solutions z^{E} and z^{M} defined in (E3) and (M3). Again, as in Theorem 4.4, this is a result in the weak sense.

Theorem 5.1. *Let $p \in [2, \infty)$ and $\mu \in \mathscr{P}(\mathbb{R})$ have the properties* (29). *Then we can define a q-dimensional standard Wiener process $(w(t))_{t\in[t_0,T]}$ and a set of i.i.d. r.v.'s $\{\xi^k_{ji}$: $j = 1, \dots, q$; $i = 1, \dots, m_k$; $k = 0, \dots, n-1\}$ with distribution $D(\xi^0_{11}) = \mu$ on a common probability space, such that for* (I) *and the methods* (E3) *and* (M3) *constructed with them we have*:

(a) *If* (V1) *and* (V2) *hold then*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - z^{\mathrm{E}}(t)\|^p \leqslant K \left\{ h^{p/2} + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \right\}.$$

(b) *If* (V1)–(V4) *hold then*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - z^{\mathrm{M}}(t)\|^p \leqslant K \left\{ h^p + \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^p \right\}.$$

Proof. To show that both assertions (a) and (b) follow from the Theorems 2.6, 3.4 and 4.4 it suffices to verify that

$$\frac{h}{m(h)} \left(1 + \ln \left(\frac{m(h)}{h} \right) \right) \leqslant K \left(\frac{1 + \ln m(h)}{\sqrt{m(h)}} \right)^2.$$

But this follows easily from (46) for $\gamma = 1/h \geqslant 1$ (because of (G1)) and $\delta = m(h) \geqslant 1$. $\square$

Since Theorem 5.1 provides results in the weak sense, it is appropriate to formulate it as an estimate for the L^p Wasserstein metric between the *distributions* of the exact solution and the approximate solutions:

Corollary 5.2. *Let $p \in [1, \infty)$ and $\mu \in \mathscr{P}(\mathbb{R})$ have the properties* (29). *Moreover, let $(w(t))_{t\in[t_0,T]}$ be a q-dimensional standard Wiener process and $\{\xi^k_{ji}$: $j = 1, \dots, q$; $i = 1, \dots, m_k$; $k = 0, \dots, n-1\}$ a set of i.i.d. r.v.'s with distribution $D(\xi^0_{11}) = \mu$. Then for* (I) *and the methods* (E3) *and* (M3) *constructed with them we have*:

(a) *If* (V1) *and* (V2) *hold then*

$$W_p(D(x), D(z^{\mathrm{E}})) \leqslant K \left\{ h^{1/2} + \frac{1 + \ln m(h)}{\sqrt{m(h)}} \right\}.$$

(b) *If* (V1)–(V4) *hold then*

$$W_p(D(x), D(z^{\mathrm{M}})) \leqslant K \left\{ h + \frac{1 + \ln m(h)}{\sqrt{m(h)}} \right\}.$$

Proof. For $p \in [2, \infty)$ the assertions follow directly from Theorem 5.1 and after applying Lemma 2.2 to the right-hand sides. Then the assertions are also true for $p \in [1, 2)$, since $W_{p_1} \leqslant W_{p_2}$ for $1 \leqslant p_1 \leqslant p_2 < \infty$ (see [7]). □

The estimates in Theorem 5.1 and Corollary 5.2 give convergence rates w.r.t. h for the methods (E3) and (M3) and for any grid sequence in $\mathscr{G}(m, \Lambda, \alpha, \beta)$. These rates consist of two summands, one depending on h and the other depending on $m(h)$, representing the rates of time and chance discretization, respectively. Obviously it is not desirable that one of both summands converges faster than the other for this would only increase the costs in relation to the effect. Namely, if the second summand converges faster than the first, this would mean that $m(h)$ increases too fast and consequently—because of (G3)—to have too small step sizes of the whole *fine* grid, i.e., to have too many points u_i^k in relation to the t_k in each grid and therefore to use a random number generator too often. If the first summand converges faster than the second, then $m(h)$ increases too slowly, i.e., the intervals $[t_k, t_{k+1}]$ do not have enough intermediate grid points u_i^k, such that the chance discretization does not keep up with the time discretization. Therefore, it is desirable to tune the rates of both summands, i.e., to equal the powers of h in both summands. This means to choose $m(h)$ to be increasing like $1/h$ for method (E3) and like $1/h^2$ for the method (M3). In this way we get the following two corollaries immediately from Theorem 5.1 and Corollary 5.2.

Corollary 5.3. *Let* $p \in [2, \infty)$ *and* $\mu \in \mathscr{P}(\mathbb{R})$ *have the properties* (29). *Then we can construct the solutions in* (I), (E3), *and* (M3) *on a common probability space* (*as in Theorem* 5.1) *such that we have*:

(a) *If* (V1) *and* (V2) *hold and* $\max\{\sup_{0<s\leqslant 1} sm(s), \sup_{0<s\leqslant 1}(1/sm(s))\} \leqslant K$ *then*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - z^{\mathrm{E}}(t)\|^p \leqslant K \cdot h^{p/2}(1 - \ln h)^p.$$

(b) *If* (V1)–(V4) *hold and* $\max\{\sup_{0<s\leqslant 1} s^2 m(s), \sup_{0<s\leqslant 1}(1/s^2 m(s))\} \leqslant K$ *then*

$$E \sup_{t_0 \leqslant t \leqslant T} \|x(t) - z^{\mathrm{M}}(t)\|^p \leqslant K \cdot h^p(1 - \ln h)^p.$$

Corollary 5.4. *Under the general assumptions in Corollary* 5.2 *we have*:

(a) *If* (V1) *and* (V2) *hold and* $\max\{\sup_{0<s\leqslant 1} sm(s), \sup_{0<s\leqslant 1}(1/sm(s))\} \leqslant K$ *then*

$$W_p(D(x), D(z^{\mathrm{E}})) \leqslant K \cdot h^{1/2}(1 - \ln h).$$

(b) *If* (V1)–(V4) *hold and* $\max\{\sup_{0<s\leqslant 1} s^2 m(s), \sup_{0<s\leqslant 1}(1/s^2 m(s))\} \leqslant K$ *then*

$$W_p(D(x), D(z^{\mathrm{M}})) \leqslant K \cdot h(1 - \ln h).$$

Thus, given a grid sequence in $\mathscr{G}(m, \Lambda, \alpha, \beta)$ with $h \to 0$ and using the metric W_p, we have under the assumptions of Corollary 5.4(a) for the method (E3) the convergence rate $\mathrm{O}(h^{1/2}(1 - \ln h))$ w.r.t. the maximal step sizes h of the coarse subgrids and the convergence rate $\mathrm{O}((h/m(h))^{1/4}(1 - \ln(h/m(h))))$ w.r.t. the maximal step sizes $h/m(h)$ of the whole fine grids and the convergence rate $\mathrm{O}(N^{-1/4}(1 + \ln N))$ w.r.t. the number N of all gridpoints of the whole fine grids. Analogously, under the assumptions of Corollary 5.4(b) we have for the method (M3) the convergence rates $\mathrm{O}(h(1 - \ln h))$, $\mathrm{O}((h/m(h))^{1/3}(1 - \ln(h/m(h))))$, and $\mathrm{O}(N^{-1/3}(1 + \ln N))$. Theorem 1.1 deals with the method (E3) in the case of $m(h) \equiv 1$ and equidistant grids, and it yields at most the convergence rate $O(N^{-1/6}(\ln N)^{\varepsilon})$ ($\varepsilon > \frac{1}{2}$). Kanagawa's result [10] does not follow from the results proved here and was proved using different tools and different assumptions. Our methods (E3) and (M3) yield better orders (essentially $N^{-1/4}$ and $N^{-1/3}$) than Kanagawa's method (K) (essentially $N^{-1/6}$). Moreover, (E3) and (M3) need to compute the coefficients b and σ only in a small part of the N grid points, namely the points t_k of the coarse subgrids, whereas (K) requires the computation of the coefficients in all N grid points. This shows that (E3) and (M3) have also lower costs than (K) for the same N. If we take in the grids for (E3), (M3), and (K) the same numbers n of "expensive" grid points (i.e., points where b and σ have to be computed) or the same corresponding maximal step sizes h, then the orders of (E3), (M3), and (K) are essentially $n^{-1/2}$, n^{-1}, and $n^{-1/6} = N^{-1/6}$ (or $h^{1/2}$, h, and $h^{1/6}$), which makes the different convergence rates more significant.

Acknowledgement

I would like to thank Werner Römisch for helpful discussions and advice while I was working on this subject.

References

[1] J.M.C. Clark and R.J. Cameron, The maximum rate of convergence of discrete approximations for stochastic differential equations, Lecture Notes in Control and Inform. Sci. **25** (Springer, Berlin, 1980) 162–171.
[2] H. Doss, Liens entre équations differentielles stochastiques et ordinaires, *Ann. Inst. Henri Poincaré* **XIII** (1977) 99–125.
[3] R.M. Dudley, Distances of probability measures and random variables, *Ann. Math. Statist.* **39** (1968) 1563–1572.
[4] M. Gelbrich, L^p-Wasserstein-Metriken und Approximationen stochastischer Differentialgleichungen, Dissertation A, Humboldt-Universität zu Berlin, Sektion Mathematik, (1989).
[5] M. Gelbrich, On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces, *Math. Nachr.* **147** (1990) 185–203.
[6] I.I. Gikhman and A.W. Skorokhod, *Introduction to the Theory of Stochastic Processes* (Nauka, Moscow, 1977, in Russian).
[7] C.R. Givens and R.M. Shortt, A class of Wasserstein metrics for probability distributions, *Michigan Math. J.* **31** (1984) 231–240.
[8] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, 1981).
[9] R. Janssen, Discretization of the Wiener-process in difference-methods for stochastic differential equations, *Stochastic Process. Appl.* **18** (1984) 361–369.
[10] S. Kanagawa, The rate of convergence for approximate solutions of stochastic differential equations, *Tokyo J. Math.* **12** (1986) 33–48.
[11] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1992).
[12] J. Komlós, P. Major and G. Tusnády, An approximation of partial sums of independent RV's, and the sample DF, I, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **32** (1975) 111–131.
[13] J. Komlós, P. Major and G. Tusnády, An approximation of partial sums of independent RV's, and the sample DF, II, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **34** (1976) 33–58.
[14] G. Maruyama, Continuous Markov processes and stochastic equations, *Rend. Circolo Math. Palermo* **4** (1955) 48–90.

[15] G.N. Milshtein, A method of second-order accuracy integration of stochastic differential equations, *Theor. Probab. Appl.* **23** (1978) 396–401.
[16] G.N. Milshtein, *Numerical Integration of Stochastic Differential Equations* (Izd. Ural. Univ. Sverdlovsk, 1988, in Russian).
[17] N.J. Newton, An asymptotically efficient difference formula for solving stochastic differential equations, *Stochastics* **19** (1986) 175–206.
[18] E. Pardoux and D. Talay, Discretization and simulation of stochastic differential equations, *Acta Appl. Math.* **3** (1985) 23–47.
[19] E. Platen, An approximation method for a class of Itô processes, *Lietuvos Mat. Rink.* XXI **1** (1981) 121–133.
[20] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models* (Wiley, New York, 1991).
[21] W. Römisch and A. Wakolbinger, On Lipschitz dependence in systems with differentiated inputs, *Math. Ann.* **272** (1985) 237–248.
[22] A.N. Shiryaev, *Probability Theory* (Springer, Berlin, 1984).
[23] R.M. Shortt, Universally measurable spaces: An invariance theorem and diverse characterizations, *Fund. Math. Th.* **121** (1983) 35–42.
[24] H. Sussmann, On the gap between deterministic and stochastic differential equations, *Ann. Probab.* **6** (1978) 19–41.
[25] D. Talay, Résolution trajectorielle et analyse numérique des équations differentielles stochastiques, *Stochastics* **9** (1983) 275–306.
[26] W. Wagner, Monte Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples, *Stochastic Anal. Appl.* **6** (1988) 447–468.